



HAL
open science

Étude de l'hétérogénéité spatiale de coupes anatomo-pathologiques par machine learning et approches statistiques : application à l'ostéosarcome

Anthony Mancini

► **To cite this version:**

Anthony Mancini. Étude de l'hétérogénéité spatiale de coupes anatomo-pathologiques par machine learning et approches statistiques : application à l'ostéosarcome. Dynamique des Fluides [physics.flu-dyn]. Institut National Polytechnique de Toulouse - INPT, 2023. Français. NNT : 2023INPT0068 . tel-04958910

HAL Id: tel-04958910

<https://theses.hal.science/tel-04958910v1>

Submitted on 20 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Dynamique des fluides

Présentée et soutenue par :

M. ANTHONY MANCINI

le vendredi 17 mars 2023

Titre :

Étude de l'hétérogénéité spatiale de coupes anatomo-pathologiques par
"machine learning" et approches statistiques : application à
l'ostéosarcome

École doctorale :

Mécanique, Énergétique, Génie civil, Procédés (MEGeP)

Unité de recherche :

Institut de Mécanique des Fluides de Toulouse (IMFT)

Directeurs de Thèse :

MME SYLVIE LORTHOIS

MME PAULINE ASSEMAT

Rapporteurs :

MME ALINE BEL-BRUNON, INSA LYON

MME MICHELE GOUFFES, UNIVERSITE PARIS-SACLAY

Membres du jury :

M. OLIVIER SAUT, CNRS AQUITAINE, Président

MME ANNE GOMEZ, CHU TOULOUSE, Membre

MME AURÉLIE DUTOUR, CENTRE LEON BERARD LYON, Membre

MME PAULINE ASSEMAT, TOULOUSE INP, Membre

MME SYLVIE LORTHOIS, TOULOUSE INP, Membre

M. PASCAL SWIDER, TOULOUSE INP, Membre

Résumé

Résumé — L'ostéosarcome (OST) est le cancer osseux primaire le plus fréquent qui touche principalement les enfants et les jeunes adultes. Les OST sont caractérisés par une production anarchique d'os par les cellules tumorales, associée à une grande complexité génomique entraînant une hétérogénéité spatio-temporelle de la tumeur et de son microenvironnement. L'hétérogénéité intra-tumorale et interindividuelle qui en résulte est explorée à différentes échelles de types génétique, moléculaire, tissulaire ou macroscopique. Le traitement de ces données multi-échelles par des techniques numériques telles que l'IA ou la modélisation basée sur l'image, dans le but de proposer des stratégies thérapeutiques personnalisées, est un domaine de recherche en plein essor.

Ce travail de thèse s'inscrit dans cette démarche en prenant pour hypothèse qu'à l'échelle du microenvironnement, le tissu tumoral peut être considéré comme un milieu poreux multiphasique réactif évoluant en espace et en temps. Cette géométrie tissulaire complexe associe des phases continues osseuse, fluide et vasculaire aux phases discrètes des populations cellulaires impliquées dans le développement tumoral et dans la réponse immunitaire. L'objectif de ces travaux est de proposer de nouvelles méthodologies statistiques ainsi que les outils numériques associés permettant, à partir d'images histologiques et immunohistochimiques de routine clinique et ayant une résolution spatiale sub-cellulaire, d'établir une quantification de l'hétérogénéité intra et inter-tumorale de l'OST à l'échelle tissulaire. Ce travail interdisciplinaire est bâti sur la collaboration entre l'Institut de Mécanique des Fluides de Toulouse et l'Institut Universitaire du Cancer de Toulouse.

L'état de l'art présente le contexte clinique et l'appropriation des techniques d'imagerie histologiques permettant de définir les questionnements scientifiques associés à cette thèse. Les méthodologies de segmentation d'images basées sur des méthodes d'apprentissage machine, sont par la suite développées. L'étape suivante du travail s'attache à la recherche de filtres linéaires et non linéaires pour l'amélioration de la segmentation obtenue par méthode d'apprentissage machine seule. Ces développements basés sur une approche analytique permettent en particulier d'interpréter de manière pertinente les filtres obtenus par des algorithmes du type Support Vector Machine. La troisième partie des travaux de recherche propose des métriques de caractérisation des images; métriques inspirées de la géostatistique. Des cartes de corrélations locales associant les propriétés de la microarchitecture poreuse et la distribution des populations cellulaires immunitaires, macrophages et lymphocytes, sont établies. La pertinence des méthodologies développées est évaluée sur une cohorte de huit patients atteints d'OST et l'hétérogénéité spatiale de la réponse au traitement est explorée. En dernier lieu, les outils numériques développés dans cette thèse sont utilisés pour l'exploration mécanobiologique de l'hétérogénéité du microenvironnement, et ceci concerne l'étude des propriétés élastiques et de transport du milieu.

En conclusion, la méthodologie développée dans cette thèse a permis de quantifier la complexité architecturale du microenvironnement tumoral en association à la distribution des cellules immunitaires permettant l'exploration de marqueurs spécifiques de résistance aux traitements. Cette approche originale à notre connaissance, permet de quantifier les phénomènes mécanobiologiques à partir d'examen cliniques établis en routine. Elle peut contribuer à terme, à l'identification de cibles thérapeutiques spécifiques aux patients.

Mots clés : Traitement d'images, machine learning, interprétabilité, modélisation basée sur images, mécanobiologie, milieux poreux, ostéosarcome, immunohistochimie, résistance au traitement

Abstract — Osteosarcoma (OST) is the most common primary bone cancer that mainly affects children and young adults. OSTs are characterized by anarchic bone production by tumor cells, associated with high genomic complexity resulting in spatiotemporal heterogeneity of the tumor and its microenvironment. The resulting intra-tumor and inter-individual heterogeneity is explored at different genetic, molecular, tissue or macroscopic scales. With the aim of proposing personalized therapeutic strategies, the processing of these multiscale data by numerical techniques such as AI or image-based modeling, is a rapidly growing research field.

This thesis is in line with this approach by assuming that at the microenvironment scale, the tumor tissue can be considered as a reactive multiphase porous medium evolving in space and time. This complex tissue geometry includes continuous bone, fluid and vascular phases and discrete phases of cell populations involved in tumor development and immune response. The objective of this work is to propose new statistical methodologies and associated numerical tools allowing, from routine clinical histological and immunohistochemical images with a sub-cellular spatial resolution, to establish a quantification of the intra and inter-tumoral heterogeneity of the OST at the tissue scale. This interdisciplinary work is built on the collaboration between the Institut de Mécanique des Fluides de Toulouse et and the Institut Universitaire du Cancer de Toulouse.

The first part presents the clinical context and the histological imaging techniques allowing to define the scientific questions associated with this thesis. Image segmentation methodologies based on machine learning methods are then developed. The next step of the work focuses on the research of linear and non-linear filters for the improvement of the segmentation obtained by machine learning method alone. These developments, based on an analytical approach, allow to interpret, in a relevant way, the filters obtained by algorithms such as Support Vector Machine. The third part of the work proposes metrics for image characterization, inspired by geostatistics. Local correlation maps associating the properties of the porous microarchitecture and the distribution of immune cell populations such as macrophages and lymphocytes are established. The relevance of the developed methodologies is evaluated on a cohort of eight patients with OST and the spatial heterogeneity of the response to treatment is explored. Finally, the numerical tools developed in this thesis are used for the mechanobiological exploration of the heterogeneity of the microenvironment, by studying the elastic and transport properties of the medium.

In conclusion, the methodology developed in this thesis has allowed to quantify the architectural complexity of the tumor microenvironment, in association with the distribution of immune cells, allowing the exploration of specific markers of resistance to treatments. This original approach, to our knowledge, also allows the quantification of mechanobiological phenomena from routine examinations, and could contribute to the identification of patient-specific therapeutic targets.

Keywords: Image processing, machine learning, interpretability, image-based modeling, mechanobiology, porous media, osteosarcoma, immunohistochemistry, resistance to treatment

Remerciements

Je remercie très chaleureusement mes directrices de thèse Pauline Assemat et Sylvie Lorthois ainsi que Anne Gomez et Pascal Swider pour leur accompagnement, leur patience, leur bonne humeur communicative et leur bienveillance.

Je remercie les chercheurs permanents, stagiaires, doctorants, post-doc, ingénieurs de recherche, techniciens et personnels de l'administration de mon laboratoire pour toutes les discussions intéressante et l'entraide à propos du vécu d'une thèse.

Je remercie le dispositif Aspie-Friendly, l'association La Bulle !, et les membres du café d'Hélène Vial pour leur soutien.

Je remercie Julien Davadant pour son aide précieuse.

Je tiens également à remercier ma famille : mes parents, dont ma mère qui a été très attentive à la manière dont oscillait mon moral tout au long de ma thèse et a tenu informé mon père. Ma soeur Sandra. Mes oncles et tantes (en incluant ceux par alliance et leur belle-famille) et mes grand-parents, dont tonton Serge qui m'avait envoyer un SMS amusant qui a su me faire rire, le matin où je faisais une conférence l'après-midi même à Oxford et pour laquelle j'étais dans un stress important.

Mes cousins Lisa, Marine, Jérémy, Danilo, Flavio et Mathis, ainsi que les filles de ma cousine Lisa : Chloé et Léa.

Contents

Résumé	i
Table des sigles et acronymes	vii
1 Introduction	1
1.1 Ostéosarcome	1
1.2 Hétérogénéité tumorale à l'échelle du micro-environnement	5
1.3 Hypothèse et approches proposées	7
2 Segmentation de coupes histologiques et immunohistologiques	11
2.1 Coupes histologiques et immunohistologiques: images cliniques étudiées	12
2.2 Généralités sur la classification et la segmentation	13
2.3 Segmentation des coupes histologiques	14
2.4 Classification en fonction de la couleur des pixels	15
2.5 Visualisation de la répartition des couleurs des pixels	15
2.6 Quelques exemples de classification des images avec des algorithmes non-supervisés et supervisés	16
2.7 Segmentation utilisant l'algorithme KNN	20
2.8 Conclusion	27
3 Recherche de filtres pour l'amélioration de la segmentation	29
3.1 Nomenclature	30
3.2 Introduction : filtres linéaires	31
3.3 Décomposition d'un filtre linéaire sur un signal "lisse".	36
3.4 Décomposition d'un filtre linéaire sur un signal bruité	43
3.5 Opérateurs différentiels invariants par rotation et symétrie	47
3.6 Variance locale	60
3.7 Conclusion	62
4 Métriques et cartes de corrélation locale	65
4.1 Porosité et définition de longueurs caractéristiques	65
4.2 Périmètre spécifique	71
4.3 Mesure locale des dimensions caractéristiques du tissu osseux	73
4.4 Calcul de la densité cellulaire	75
4.5 Calcul du facteur de pertinence	78
4.6 Corrélations entre quantificateurs	78
5 Application des algorithmes sur une cohorte de patients	93
5.1 Correlations globales	93
5.2 Patients mauvais répondeurs: patient 2	94
5.3 Patient bon répondeur: patient 7	97
5.4 Conclusion	99
6 Vers l'imagerie fonctionnelle microscopique numérique	101
6.1 Effets mécaniques et ostéosarcome	101
6.2 Méthodologie pour l'imagerie fonctionnelle numérique	102

Conclusion	107
A Annexe 1: Recalage spatial	111
A.1 Recalage spatial	111
B Annexe 2: Coefficient de Spearman local	119
C Annexe 3: Résultats complémentaires sur la cohorte de patients	121
C.1 Patients mauvais répondeurs	121
C.2 Patients bons répondeurs	125
Bibliographie	139

Table des sigles et acronymes

[OS] Ostéosarcome

[MEC] Matrice Extra Cellulaire

[OC] Ostéoclastes

[OMS] Organisation Mondiale de la Santé

[H&E] Hémalum-éosine

[CD8, CD163, CD68, CD34] Marquages d'immuno-histologie, CD8 : lymphocytes T, CD163 : macrophages plutôt de bon pronostic, CD68 : macrophages et ostéoclastes plutôt de mauvais pronostic, CD34 : cellule endothéliales des vaisseaux sanguins.

[CD68t, CD68g] CD68 cellules Totales, CD68 Grandes cellules

[BR] Bon Répondeur (de la tumeur à la chimio-thérapie)

[MR] Mauvais Répondeur (de la tumeur à la chimio-thérapie)

[IA] Intelligence Artificielle

[KNN] k-Nearest-Neighbors, k plus proches voisins

[RF] Random Forest

[CPU] Central Processing Unit

[SVM] Support Vector Machine

[px] pixels

[MEF] Méthode des Éléments Finis

Chapter 1

Introduction

Contents

1.1 Ostéosarcome	1
1.1.1 Ostéosarcome: clinique	1
1.1.2 Ostéosarcome: biologie	3
1.2 Hétérogénéité tumorale à l'échelle du micro-environnement	5
1.3 Hypothèse et approches proposées	7

Cette thèse propose une méthode d'analyse spatiale pour explorer l'hétérogénéité du microenvironnement tumoral, à partir d'images histologiques et immuno-histochimiques cliniques, dans le contexte de l'étude de l'ostéosarcome. Ce chapitre d'introduction présente dans un premier temps, une courte synthèse de la littérature concernant la compréhension actuelle de l'ostéosarcome d'un point de vue clinique et biologique. Par la suite, une description de l'état de l'art et des enjeux concernant l'étude de l'hétérogénéité du microenvironnement tumoral est développée. Les questionnements et approches proposés seront détaillés dans la dernière section de ce chapitre.

1.1 Ostéosarcome

1.1.1 Ostéosarcome: clinique

L'ostéosarcome (OST) est une tumeur osseuse maligne primaire rare (350 nouveaux cas/an en France) qui touche principalement les enfants et les jeunes adultes de moins de 25 ans [1]. Il s'agit du cancer osseux le plus fréquent, caractérisé par la production anarchique de matrice osseuse par les cellules tumorales [2]. L'OST survient dans les os à croissance rapide des patients tels que le fémur, le tibia, l'humérus, principalement autour des plaques de croissance. À ce jour, le traitement de l'OST consiste le plus souvent en des chimiothérapies néoadjuvantes ¹ suivies d'une résection chirurgicale conservatrice de l'os affecté, et en des chimiothérapies post-chirurgie adaptées à la réponse au traitement initial [3]. L'évaluation initiale de la réponse au traitement est réalisée par cartographie de la réponse, exprimée en pourcentage de cellules viables, sur une section totale de la résection chirurgicale de la tumeur primitive. Les patients sont classés comme "bon répondeur" ou "mauvais répondeur" au traitement, en fonction du pourcentage de cellules tumorales viables (ou non nécrosées), avec un seuil fixé à 10% et ceci selon 4 grades. Ainsi, les patients présentant un pourcentage de cellules tumorales nécrosées inférieur à 90% sont classés comme mauvais répondeurs, et les autres, comme bons répondeurs ([4], table 1.1).

¹Cycles de chimiothérapie précédant la chirurgie

	Grade Huvos et Rosen	Chimiothérapie	Pourcentage de cellules tumorales nécrosées
Mauvais Répondeur	Grade I	Pas d'effet	< 10%
Mauvais Répondeur	Grade II	Peu efficace	10 – 90%
Bon Répondeur	Grade III	Moyennement efficace	91 – 99%
Bon Répondeur	Grade IV	Très efficace	> 99%

Table 1.1: Classification de la réponse au traitement par chimiothérapie néoadjuvante chez les patients atteints d'ostéosarcome selon les grades de Huvos et Rosen, et critères associés.

L'OST se distingue des autres sarcomes par la production par les cellules tumorales d'une matrice extra-cellulaire de type osseux, et d'étendue variable. On retrouve de fines lames ostéoïdes ou épaisses travées ostéochondroblastiques. Dans la dernière classification de l'OMS 2020, les OST conventionnels de haut grade sont ainsi subdivisés en 3 sous-types : les formes ostéoblastiques, les plus fréquentes, les formes chondroblastiques et les formes fibroblastiques [1] tels qu'illustrées par leurs coupes histologiques en coloration H&E (Hemalun et éosine) dans la figure 1.1. Une des caractéristiques de l'OST est qu'elle présente une grande complexité génomique [5, 6, 7], responsable d'une hétérogénéité intratumorale, identifiable à différentes échelles (par exemple au niveau génétique, métabolique, moléculaire, microscopique tissulaire et cellulaire), variable d'un individu à l'autre, et évoluant dans le temps par l'acquisition d'anomalies moléculaires supplémentaires.

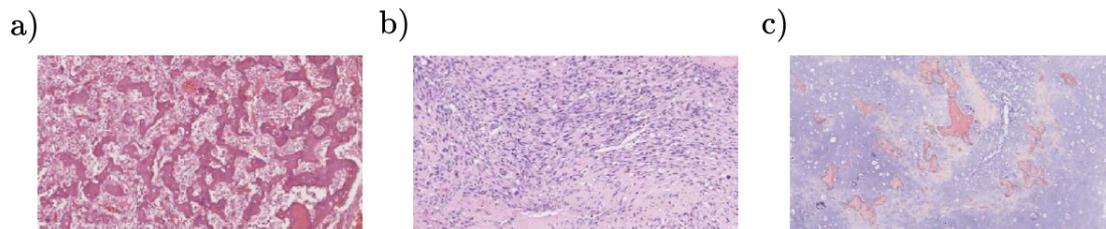


Figure 1.1: Coupes histologiques (coloration H&E) des trois catégories conventionnelles a) ostéoblastique, b) fibroblastique, c) chondroblastique, Centre de Ressources Biologiques - Cancer de Toulouse.

La coloration à l'éosine-hemalun H&E (figure 1.2 a) marque principalement le collagène en rose, le noyau des cellules en violet et les tissus cartilagineux en bleu. L'immunohistologie est une autre technique de coloration (figure 1.2 b) qui marque en marron certains éléments possédant un antigène par fixation d'anticorps puis de l'espèce chimique colorante. Ainsi, certaines cellules du système immunitaire comme les macrophages, lymphocytes, ou les parois des vaisseaux sanguins (cellules endothéliales) peuvent être marquées par immunohistochimie (voir chapitre 2).

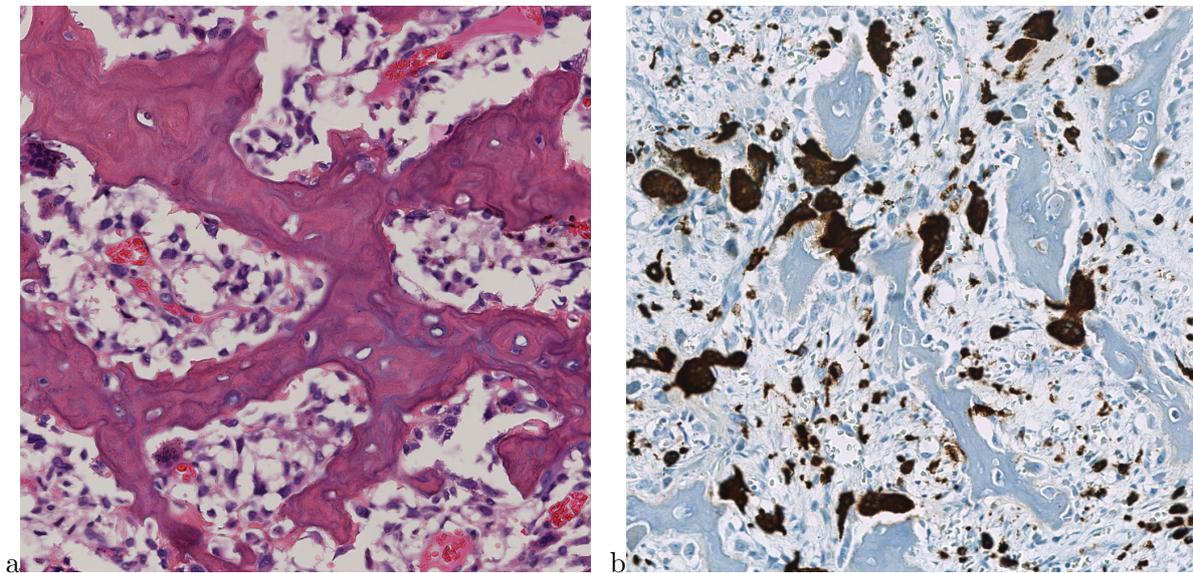


Figure 1.2: a) Coupe histologique marquée à l'hémalum-éosine H&E, b) coupe immunohistologique (marquage CD68). Ces marquages permettent de mettre en évidence les travées osseuses (a : rose/violet, b : bleu), les noyau des cellules (a : violet, b : bleu pour les cellules non marquées et marron pour les cellules marquées), les fibres (a : rose; b : bleu clair). L'immunohistologie marque en marron des populations cellulaires cibles. Ces éléments sont indiqués plus précisément dans la figure 1.3

1.1.2 Ostéosarcome: biologie

L'étude biologique de l'OST est un problème extrêmement complexe en raison de la nature multi-échelle de ces tumeurs d'un point de vue spatio-temporel, et en raison de son hétérogénéité à différentes échelles. Chacune de ces échelles est généralement étudiée indépendamment et sans lien quantitatif multi-échelle. Toutefois, nous assistons à l'émergence d'approches multimodales et multi-échelles, qui en se développant, pourraient contribuer à la médecine personnalisée.

L'échelle génomique

Les techniques de séquençage de dernière génération ont confirmé le chaos extrême de l'OST, l'attribuant dans plus de 90% des cas à l'apparition d'altérations chromosomiques massives et complexes au cours d'un événement moléculaire cataclysmique unique. Récemment, l'analyse évolutive des profils transcriptomiques sur des cellules tumorales isolées (scRNAseq) provenant de quelques cas de patients atteints d'OST (<10 cas) offre de nouvelles perspectives pour comprendre l'oncogenèse de l'OST [8, 9]. Ce large chaos génomique de l'OST pourrait également être maintenu par des interactions entre les cellules tumorales et celles du microenvironnement [10].

L'échelle microscopique des tissus : le microenvironnement et l'OST

L'étude du microenvironnement de l'OST est au cœur de cette thèse. Ce microenvironnement peut être étudié à l'échelle du tissu, qui est le niveau du diagnostic microscopique par le pathologiste.

A l'échelle du micro-environnement, on peut mettre en évidence des interactions à courte distance telles que cellule/matrice extra-cellulaire, cellule/cellule et des interactions à longue distance via le transport de ces espèces solubles dans les réseaux vasculaires, lymphatiques et interstitiels. La complexité du micro-environnement découle directement de la géométrie et de la présence des différentes phases (os, fluide et cellules), des objets biologiques et de la diversité des populations cellulaires tels qu'en partie illustrés dans la figure 1.3. Dans l'OST, la biologie de la matrice extracellulaire minéralisée, non minéralisée et fibreuse, les vaisseaux, les populations cellulaires (tumorales, immunitaires et autres) et les espèces solubles, c'est-à-dire les protéines, les facteurs de croissance, les métabolites et l'oxygène, sont imbriqués. La complexité s'intensifie

lorsque leurs couplages respectifs sont explorés (figure 1.4).

À ce jour, le rôle du microenvironnement tumoral dans l'OST n'est pas élucidé. La caractérisation du microenvironnement immunitaire de l'OST par des approches immunohistochimiques et de séquençage de l'ARN a révélé que le microenvironnement immunitaire de l'OST était enrichi en cellules myéloïdes (monocytes, TAMs, cellules dendritiques, neutrophiles), et que le compartiment lymphoïde était principalement pauvre en TILs (CD8+, CD4, T-reg, et lymphocytes cytotoxiques NK) [9, 11, 12, 13]. Parmi les cellules immunitaires, toutes les données de la littérature s'accordent sur le rôle délétère des ostéoclastes [11, 12, 13, 14, 15, 16]. Les cellules souches mésenchymateuses, en se différenciant en ostéoblastes ou en chondroblastes au cours du processus oncogène, influencent la composition de MEC osseuse et du microenvironnement [17], et peuvent constituer une barrière au traitement [18, 19, 20]. Dans l'OST, la MEC est composée de manière caractéristique de phases osseuses néoformées minéralisées et non minéralisées, dont il a été démontré qu'elles contribuent à la propagation métastatique des cellules ostéosarcomateuses en influençant les populations cellulaires via leur communication ou leur adhésion [20]. L'angiogenèse est également essentielle à la croissance tumorale et à la propagation métastatique par l'apport de nutriments et d'oxygène, mais aussi par le transport propagatif des cellules cancéreuses [21].

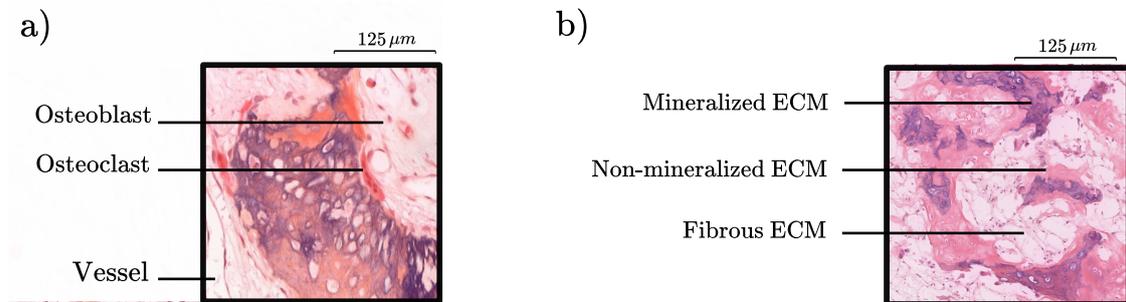


Figure 1.3: Illustration des principales phases rencontrées dans les coupes histologiques d'OST. Image Centre Ressources Biologiques -Cancer Toulouse.

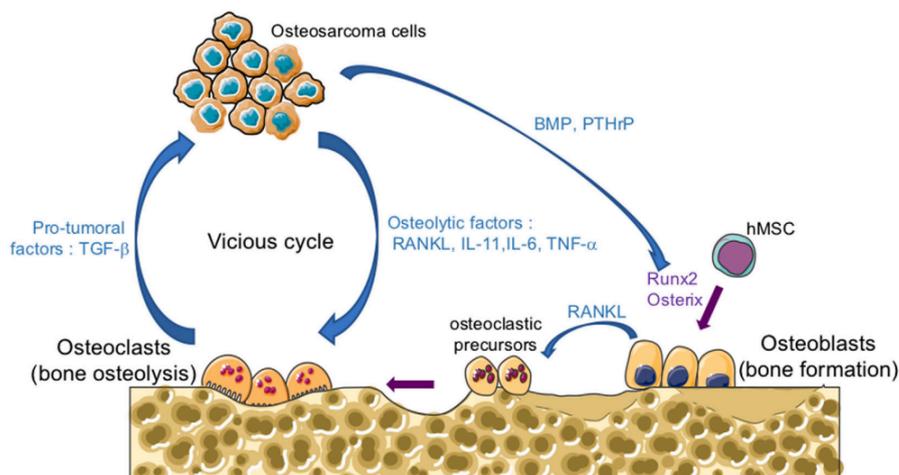


Figure 1.4: Cercle vicieux de l'ostéosarcome associant les cellules du remodelage osseux et les cellules tumorales [22].

D'un point de vue thérapeutique, la réponse des cellules tumorales au traitement est multifactorielle. Elle dépend de la biologie intrinsèque des tumeurs, mais aussi de leurs interactions dynamiques, des contraintes mécaniques, de la microcirculation tumorale, des cellules du microenvironnement immunitaire et du remodelage osseux. L'étude de l'hétérogénéité de l'ostéosarcome, et en particulier l'étude de la distribution des cellules immunitaires [23] en lien avec l'hétérogénéité de la réponse au traitement peut venir en complément des études biologiques et cliniques existantes, avec un apport original et quantitatif.

1.2 Hétérogénéité tumorale à l'échelle du micro-environnement

Le cancer est associé à des phénomènes se produisant à différentes échelles temporelles et spatiales [24] (figure 1.5), telles que la génétique, le métabolisme, la cellule, le tissu, l'organe, les individus, la cohorte et l'environnement externe. Parce qu'il est hétérogène à ces différentes échelles [25], la pathologie numérique (digital pathology) s'est avérée utile pour étudier cette hétérogénéité [26, 27, 28, 29] car elle relie la structure et la biologie des tissus, grâce à la labellisation des cellules. Par conséquent, cette approche soutenue par l'avènement de l'intelligence artificielle, est un domaine émergent [30, 31, 32]. La motivation clinique sous-jacente à l'utilisation de ces techniques est notamment que la plupart des tumeurs à génomique complexe développent, au cours de leur croissance, une hétérogénéité spatiale [33]. Ceci rend le diagnostic basé sur des biopsies standards potentiellement biaisé et les traitements cliniques efficaces uniquement dans certaines zones de la tumeur.

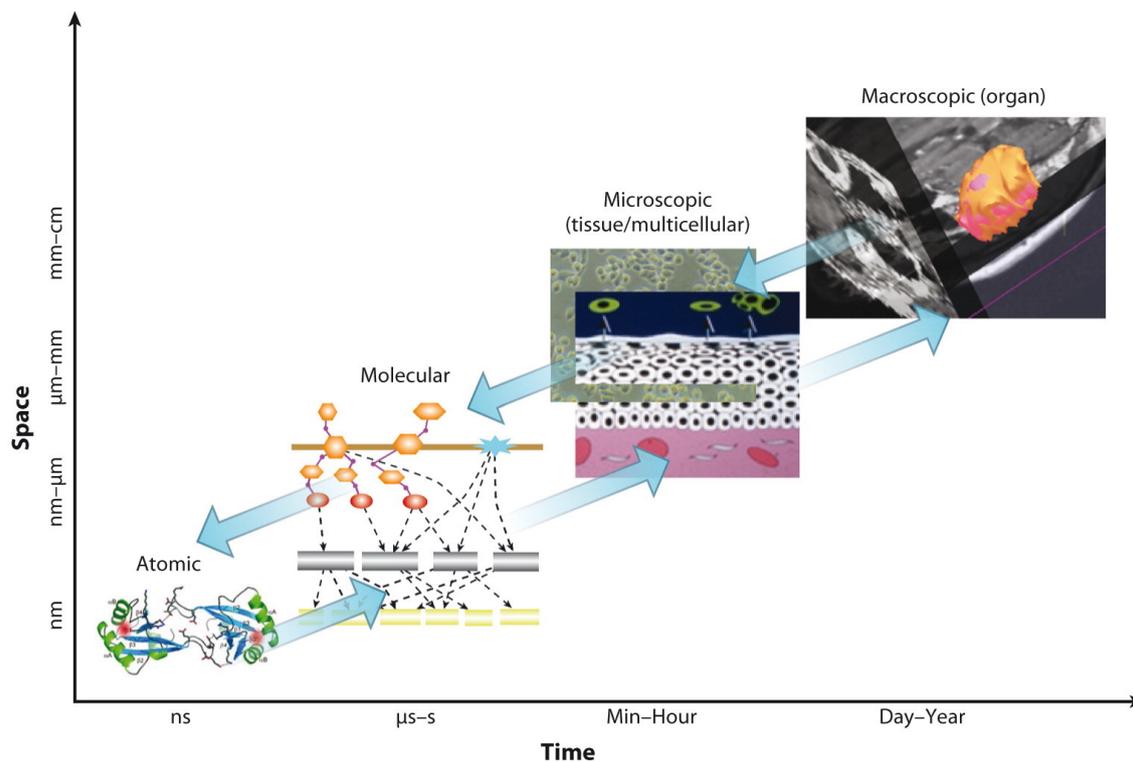


Figure 1.5: Illustration schématique des différentes échelles biologiques dans la modélisation du cancer [24].

Dans ce contexte, l'échelle de l'hétérogénéité intra-tumorale, c'est-à-dire l'échelle du micro-environnement, est d'un intérêt particulier car elle met en évidence les mécanismes, de l'échelle cellulaire à l'échelle du tissu, impliqués dans le développement du cancer, la réponse immunitaire et la résistance aux traitements [34]. Ainsi, l'étude de l'hétérogénéité intra-tumorale avec une résolution sub-cellulaire peut dévoiler de nouveaux marqueurs et de nouvelles stratégies thérapeutiques [35]. Actuellement, alors que de nombreux développements théoriques et numériques se concentrent sur l'hétérogénéité spatiale des distributions relatives des populations de cellules [27, 36, 37, 38, 39, 40, 41] illustré par exemple figure 1.6 principalement basés sur l'analyse statistique des modèles spatiaux [42], à notre connaissance, il y a un manque de techniques pour explorer l'hétérogénéité entre les cellules et la MEC alors que celle-ci joue un rôle crucial dans le développement du cancer et la résistance aux traitements [17].

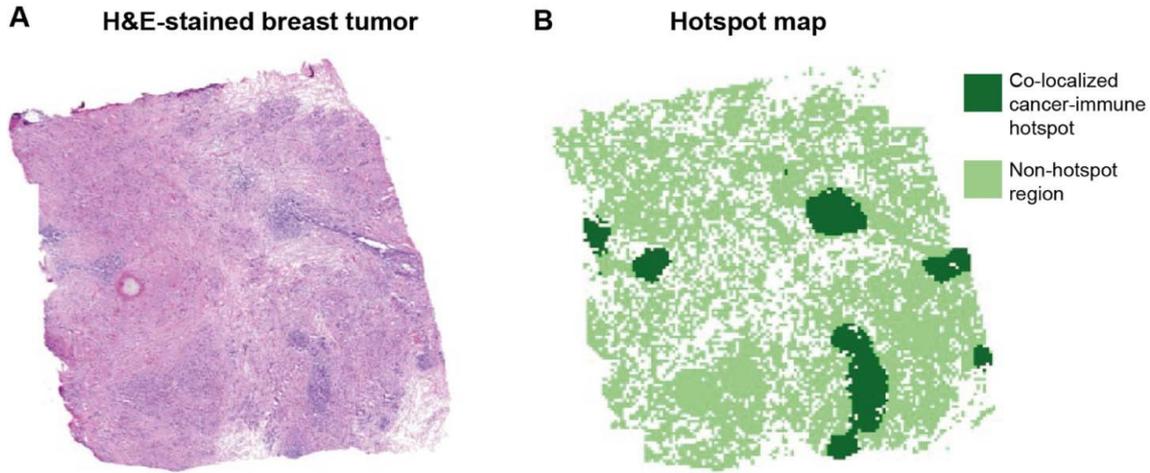


Figure 1.6: Etude de la répartition spatiale des cellules sur une coupe H&E avec analyse de points chauds par méthode de détection de Getis-Ord [27].

L'objectif premier de ce travail est de développer une méthodologie, inspirée des approches des milieux poreux, pour étudier des marqueurs structurels de l'hétérogénéité spatiale intratumorale incluant les phases cellulaires et la MEC.

Les développements proposés, sont inspirés de la géophysique et visent à calculer des corrélations entre des objets de dimensions caractéristiques différentes. En effet, la phase cellulaire est composée d'objets discrets de dimensions caractéristiques allant de 10 à $50\mu\text{m}$ alors que la phase osseuse tumorale est un ensemble continu, potentiellement anisotrope, de taille caractéristique allant de $50\mu\text{m}$ à quelques millimètres. Dans la littérature, des corrélations morphométriques entre les champs de phases continues et discrètes peuvent être examinées par deux stratégies principales : les techniques de morphologie mathématique [43, 44, 45, 46] ou les corrélations sur des champs continus par morceaux [47] (pavage). A l'heure actuelle, et à notre connaissance, aucune de ces méthodes n'a été développée dans le contexte des milieux biologiques. Elles ne peuvent d'ailleurs pas être utilisées en l'état au sens où, dans les milieux réels hétérogènes géophysiques, les recherches sont le plus souvent basées sur des données non accessibles à l'observateur, qui doivent être inférées à partir des connaissances statistiques alors que dans le contexte des données biologiques sur de grandes coupes de ressection, les données sont complexes mais disponibles. Elles nécessitent donc le développement de méthodologies adaptées.

En raison des hypothèses restrictives sur la distribution des densités de probabilité des champs dans l'approche de morphométrie mathématique (principalement gaussiennes dans les méthodes proposées dans la littérature), l'approche par pavage sera développée dans le contexte de l'exploration de tissus hautement hétérogènes dans l'espace. L'un des inconvénients de cette dernière est que, en l'absence de séparation d'échelle [48] dans de tels milieux poreux, aucun principe fondamental de convergence n'est atteint. Des tests numériques doivent alors être réalisés pour vérifier la validité de la méthode sur des pavages de différentes tailles caractéristiques. Cependant, cette deuxième stratégie, initialement développée pour étudier le transport et l'élasticité dans les milieux poreux, doit être adaptée ici pour explorer des morphométries complexes à géométrie réelle. Cette approche est en effet intéressante à explorer au vu de sa pertinence sur l'étude des champs de déformation et de transport, les données résultantes étant faciles à interpréter et pouvant être utilisées par des chercheurs et praticiens non familiers avec la théorie des milieux poreux. Elles ouvrent également la voie à l'exploration des propriétés physiques [47] de tumeurs hautement hétérogènes comme cela sera introduit dans le chapitre 6 de ce document et au développement de modèles spatio-temporels pertinents [49].

1.3 Hypothèse et approches proposées

Ainsi, parce que les tumeurs, à l'échelle des tissus, présentent plusieurs phases (fluide, cellulaire, structures molles et finalement structures rigides), l'hypothèse centrale de la méthodologie proposée dans la thèse, est qu'à l'échelle du micro-environnement, le tissu tumoral peut être considéré comme un milieu poreux multiphasique réactif, évoluant en temps et en espace.

Les images histologiques et immunohistologiques, obtenues en routine clinique pour les patients atteints d'ostéosarcome, renferment un grand nombre d'informations (biologiques et structurales) souvent peu exploitées du fait de leur grande taille pouvant atteindre 10^{10} pixels et de la complexité et diversité des données.

Alors que des techniques automatiques et optimisées sont disponibles dans la littérature pour la segmentation à travers plusieurs logiciels libres (voir une liste dans [30]), la classification du tissu osseux relativement aux cellules est difficile à réaliser dans les coupes d'ostéosarcome [50], notamment dans les coupes immunohistologiques. Nous avons donc développé une méthodologie de segmentation adaptée aux images cliniques étudiées, qui sera d'abord détaillée dans le chapitre 2 de cette thèse. À partir de ces développements numériques, potentiellement coûteux en temps de calcul, des stratégies de développements de filtres ont été proposées pour améliorer la segmentation. De nouveaux développements ont conduit à l'élaboration d'une théorie d'interprétabilité des filtres obtenus par méthodes de machine learning du type Support Vector Machine (SVM), qui constitue l'essentiel du chapitre 3 de cette thèse. Dans le chapitre 4, nous présenterons les quantités d'intérêt ou paramètres équivalents utilisés pour décrire les distributions spatiales des phases cellulaire et osseuse dans la tumeur. Nous présenterons le cadre théorique utilisé pour obtenir des cartes de corrélation spatiales d'intérêt pour les pathologistes.

Le schéma 1.7 résume l'approche proposée dans les chapitres 2 à 4. Le but est d'étudier l'hétérogénéité de la tumeur par la donnée des coupes histologiques et immunohistologiques, or étant donné la qualité plus ou moins bonnes de certaines images (déchirure des tissus, replis, flou, présence de bulles d'air), un masque est d'abord appliqué afin d'éliminer les parties qui ne peuvent pas être étudiées. Ce masque est obtenu manuellement ou par recherche de la plus grande région connexe de vide après segmentation. Du fait de la très grandes tailles des images, souvent de l'ordre de $100\,000$ pixels \times $100\,000$ pixels, le choix des algorithmes de segmentation est contraint par la nécessité d'obtenir des résultats en temps acceptable. La segmentation des images se fait en deux étapes successives, l'une utilisant le KNN (avec échantillonnage) permettant une première segmentation (chapitre 2), la suivante utilise les filtres linéaires ou non pour améliorer la qualité de la segmentation (chapitre 3). À partir de ces segmentations et après post-traitements, les populations cellulaires et le tissu osseux sont extraits et des quantités globales et locales (sous forme de cartes) peuvent être obtenues. Des coefficients de corrélation globaux et locaux sont établis (chapitre 4). Le chapitre 5 illustrera l'utilisation des développements précédents sur une cohorte de patients, appliqués sur les marquages immunohistologiques CD8, CD163 et CD68 (plusieurs populations cellulaires peuvent être distinguées dans le cadre du marquage CD68). Des résultats globaux et locaux telles que les cartes de porosité et densité cellulaire ainsi que de corrélations locales y sont illustrées. Ces résultats mettent en évidence, par méthodes numériques, de nouveaux marqueurs de la réponse au traitement. L'annexe A montre un exemple d'utilisation de recalage spatial afin de comparer différentes lames avec différents marquages tel qu'illustré entre les marquages H&E et CD8. L'annexe C propose des résultats supplémentaires. Le chapitre 6 introduira l'utilisation de ces développements pour une caractérisation fonctionnelle des ostéosarcomes, c'est à dire des propriétés de transport et élastiques, à partir des images histologiques et immunohistologiques par une approche illustrée figure 1.8. En dernier lieu, une conclusion générale sera présentée, prolongée par les perspectives de recherche.

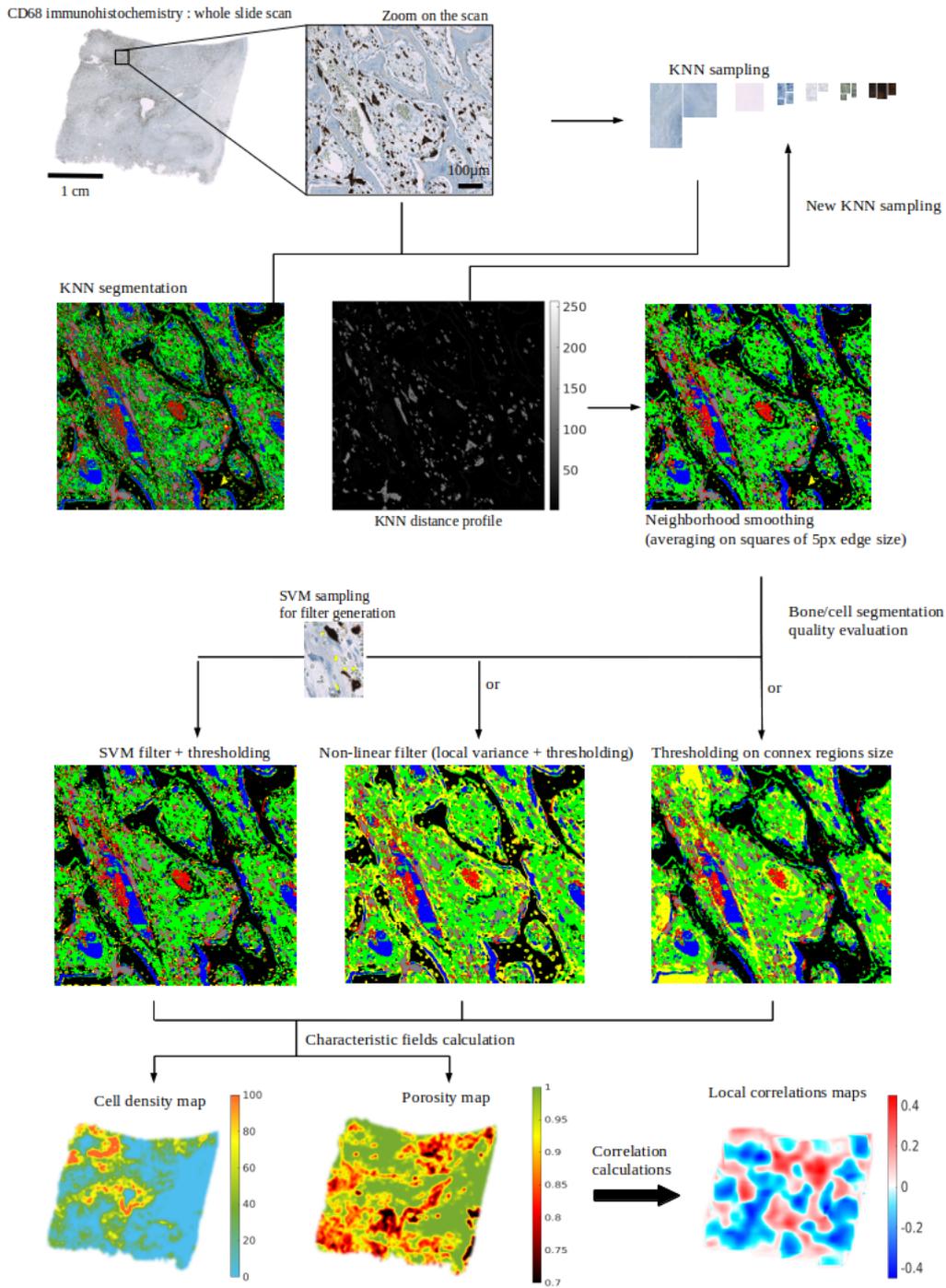


Figure 1.7: Illustration des techniques de segmentation développées dans cette thèse et de calcul des cartes de porosité et densités cellulaires pour une coupe immunohistologique marquée au CD68 (macrophages), code MECHISTO®. Lame colorée par IHC (CD68) avec une taille de pixel = $0,466\mu\text{m} \times 0,466\mu\text{m}$. Echantillonnage. Segmentation illustrée sur la région zoomée avec carte de distance. Différents filtres développés. Densité cellulaire (nombre de cellules/ mm^2 , à gauche) et porosité (au milieu). Carte de corrélation locale entre la densité cellulaire et la porosité (à droite).

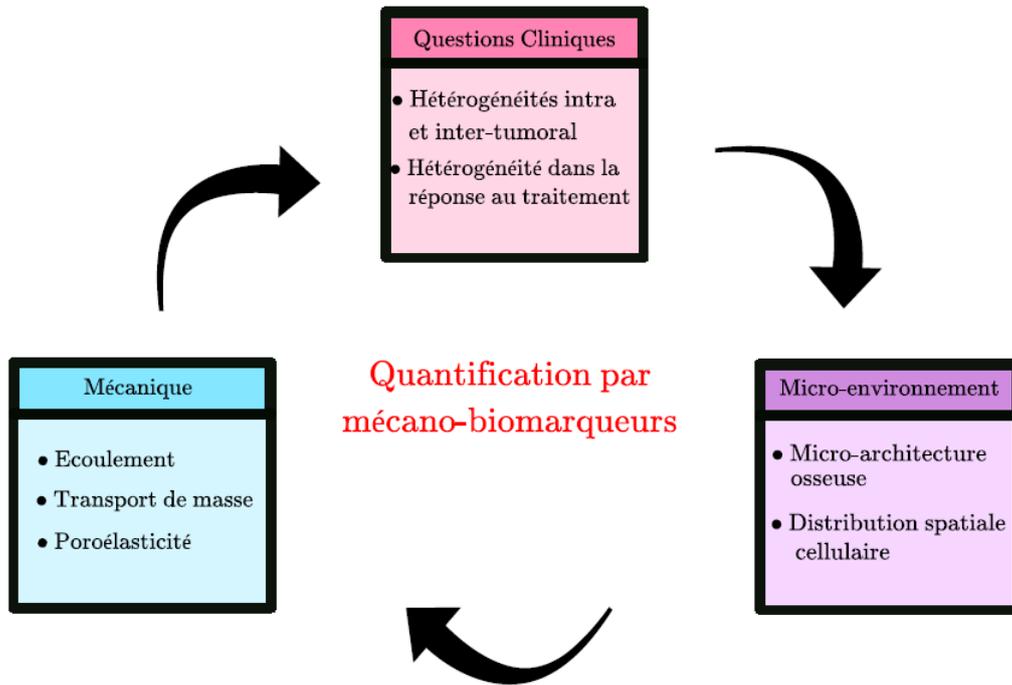


Figure 1.8: Schéma illustratif de la recherche de mécano-biomarqueurs à laquelle les développements de cette thèse participent.

Chapter 2

Segmentation de coupes histologiques et immunohistologiques

Contents

2.1	Coupes histologiques et immunohistologiques: images cliniques étudiées	12
2.2	Généralités sur la classification et la segmentation	13
2.3	Segmentation des coupes histologiques	14
2.4	Classification en fonction de la couleur des pixels	15
2.5	Visualisation de la répartition des couleurs des pixels	15
2.6	Quelques exemples de classification des images avec des algorithmes non-supervisés et supervisés	16
2.7	Segmentation utilisant l’algorithme KNN	20
2.8	Conclusion	27

Dans la première partie, la segmentation des images cliniques permettant d’extraire les phases du tissu biologique est présentée. Dans cette partie, plusieurs algorithmes ont été testés sur des images numérisées de coupes histologiques et immunohistologiques. Il est important de noter que certains logiciels libres comme Cellprofiler [51, 52], Qupath [53], Ilastik [54, 55], Icy [56] proposent des algorithmes de segmentation de coupes histologiques mais ceux qui ont été testés n’ont pas donné de résultats satisfaisants. Ces images microscopiques mettent en évidence plusieurs phases superposées et il est souvent difficile de différencier la phase osseuse de la phase cellulaire sans des algorithmes spécifiques. En effet, les ostéoblastes étant des cellules parfois bordantes de la zone osseuse, il est souvent difficile de les différencier les unes de l’autre. Ce problème est d’autant plus exacerbé pour les coupes immunohistologiques. Or, le tissu osseux et les populations cellulaires sont les phases d’intérêt pour explorer le lien entre ostéof ormation et ostéolyse, réponse immunitaire et résistance au traitement. Dans ce chapitre, nous cherchons à développer des méthodes avec une précision de plus de 95%, caractérisée par l’indice de Jaccard.

Dans la suite, la méthodologie pour obtenir ces images (cancéropôle de Toulouse) est décrite, et les éléments généraux sur la segmentation par méthode de classification par la couleur rappelés. Les différentes approches utilisées pour obtenir une méthode de segmentation satisfaisante sont ensuite exposés. Le livre d’Hastie [57] constitue une base de données importante des algorithmes existants pouvant être utilisés pour la segmentation dont certains sont présentés dans ce chapitre de thèse. Les algorithmes développés dans la thèse ont été programmés à partir du langage compilé matlab© en s’appuyant sur certaines fonctions et bibliothèques disponibles dans cet environnement.

2.1 Coupes histologiques et immunohistologiques: images cliniques étudiées

Lorsqu'un patient présente des symptômes de l'ostéosarcome, après une évaluation par imagerie médicale (scanner et/ou IRM), une biopsie de tissu tumoral, c'est-à-dire le prélèvement d'un petit volume de tissu, est réalisée. L'anatomo-pathologiste effectue le diagnostic après observation du tissu au microscope.

L'ostéosarcome conventionnel peut adopter des morphologies histologiques variées [58]. L'identification d'une matrice ostéoïde (matrice osseuse non minéralisée immature) liée au développement tumoral est essentielle au diagnostic. La tumeur prolifère fréquemment dans la cavité médullaire, détruit les travées osseuses et s'étend jusqu'à la corticale en infiltrant les canaux de Havers. Les cellules tumorales possèdent souvent des atypies nucléo-cytoplasmiques. L'activité mitotique est habituellement augmentée, avec de nombreuses mitoses atypiques. La substance ostéoïde est éosinophile, d'abondance variable en fonction du grade ou du sous type histologique. Dans l'ostéosarcome conventionnel, elle se présente sous la forme d'un réseau grêle, dit en dentelle, disposé entre les cellules tumorales mais peut également prendre la forme de travées.

La pièce de résection tumorale extraite chirurgicalement est ensuite découpée en sections de 4mm d'épaisseur (figure 2.1a-b), puis chaque coupe est fixée dans du formol à 4% pendant 48 heures [59]. Plusieurs cycles discontinus de décalcification de 4 à 6 heures suivis d'une fixation sont réalisés jusqu'au lendemain. Après décalcification complète, la pièce est divisée en 2 ou 4, en fonction de sa taille, et placée dans des cassettes adaptées. Au préalable, les prélèvements sont rincés à l'eau, avant le cycle d'imprégnation en paraffine. La coupe (4 μm d'épaisseur) est réalisée à l'aide d'un microtome à mâchoire spécifique (2.1c). La coloration de la lame est réalisée par un automate de coloration. Par exemple, la coloration H&E (hématoxyline & éosine) permet d'observer le collagène, présent dans l'os et l'ostéoïde, en rose et les noyaux des cellules en violet (2.1d). Les colorations en immuno-histologie utilisent des anticorps, modifiés de manière à donner une coloration marron, qui vont cibler des protéines de surfaces ou du cytoplasme de cellules cibles (2.1e), en particulier de cellules du système immunitaire (macrophages, lymphocytes, cellules dendritiques ...), en fonction de l'anticorps utilisé.

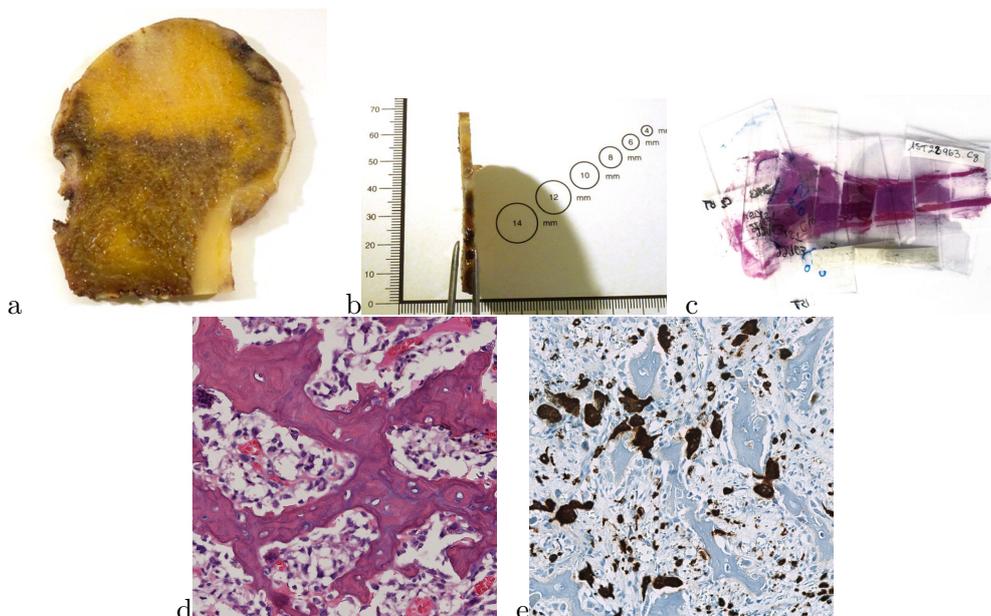


Figure 2.1: **a-c)** Extraction de coupes histologiques à partir de la pièce de résection, marquage H&E [59] **d)** scan numérique des marquages H&E, et **e)** marquage CD68 (macrophages), CRB, Cancéropole de Toulouse.

L'évaluation de la réponse à la chimiothérapie est faite par analyse microscopique des

différents prélèvements en grille effectués sur les coupes de 4mm d'épaisseur [4]. Le pourcentage de cellules nécrosées est calculé sur chacun des prélèvements. Une moyenne arithmétique des différents pourcentages obtenus sur chacun des prélèvements est ensuite réalisée et permet de définir le grade de Huvos et Rosen [60] (tableau 1.1) et une stratégie thérapeutique spécifique est adoptée. Afin d'être étudiées numériquement, ces coupes sont scannées (NanoZoomer-XR, Hamamatsu), générant des très grandes images permettant l'examen des populations cellulaires à l'échelle du tissu.

A ce jour, l'approche de l'anatomopathologiste reste qualitative. Les développements proposés dans la suite de cette thèse visent à apporter des données quantitatives qui pourront être étudiées statistiquement. La première étape est donc d'effectuer une segmentation des images. Les marquages explorés sont le H&E utilisé de manière systématique en routine clinique, le marquage CD68 qui concerne les macrophages et les ostéoclastes généralement associés à un mauvais pronostic, le marquage CD163 associé aux macrophages plutôt de bon pronostic et le marquage CD8 associé aux lymphocytes T [11, 16].

Dans la suite, les images étudiées seront traitées par des algorithmes de machine learning, pertinents dans le contexte de la segmentation d'images histologiques et immunohistologiques [54]. Étant donné la très grande taille des images en mémoire et en dimension, elles ne sont pas directement exploitables, elles sont dans un premier temps découpées en mosaïques d'images de taille 8000px×8000px. Ensuite se rajoute la contrainte d'établir des algorithmes de segmentation dont le temps d'exécution reste acceptable, il y a donc un compromis entre la complexité temporelle des algorithmes et la qualité de la segmentation. Ce compromis est délicat car il la parallélisation pourrait permettre de se permettre l'utilisation d'algorithmes gourmands en temps et ressource, mais l'outil développé dans cette thèse pourrait déboucher sur un logiciel que nous voudrions utilisable dans des hopitaux quelque soit leur localisation dans le monde sans leur imposer des ordinateurs onéreux.

2.2 Généralités sur la classification et la segmentation

En machine learning, la classification consiste à estimer un label sur un jeu de données. Soient les observations

$$(X_i, Y_i) \text{ où } Y_i \in [1, n_{\text{labels}}] \text{ est un label sur } X_i \quad \text{et } [1, n_{\text{labels}}] \text{ désigne } [1, n_{\text{labels}}] \cap \mathbb{Z} \quad (2.1)$$

tel que n_{labels} est le nombre de labels.

Ne connaissant que X_i nous essayons d'estimer Y_i , la nature de la phase, avec un estimateur $\hat{f}(X_i)$ [57].

La figure 2.2 montre un exemple de classification obtenu par l'estimateur $\hat{f}(X_i)$ de la méthode des K plus proches voisins (KNN) en vert et magenta, à partir d'observations X_i labellisées en bleu et rouge.

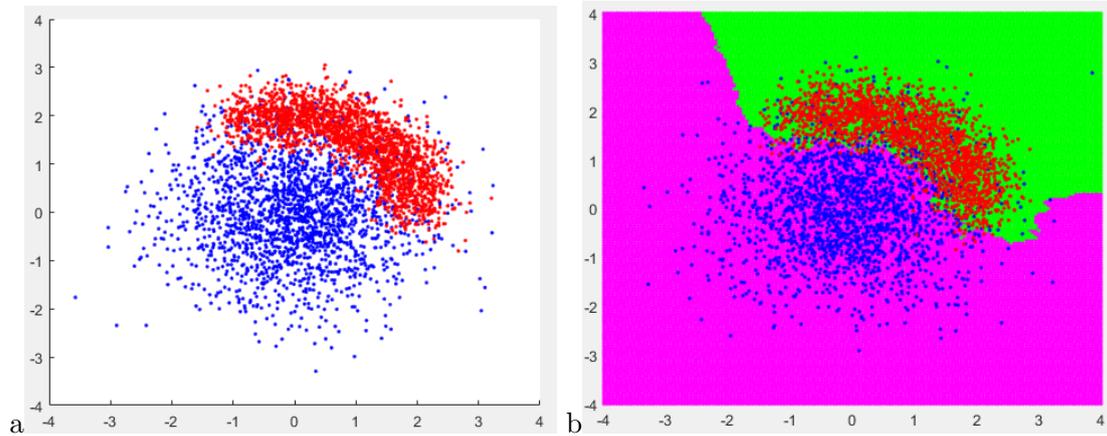


Figure 2.2: a) Exemple d’observations dans le plan ayant un label bleu ou rouge, b) classification obtenue par l’algorithme du KNN dans le plan. Une nouvelle observation se trouvant dans la partie verte (resp. magenta) du plan sera classée avec un label rouge (resp. bleu).

Lorsque la phase d’apprentissage de la classification s’appuie sur des données d’entraînement pour lesquels on connaît déjà un label, la classification est supervisée. Dans le cas contraire, l’algorithme est non supervisé, et le mot clé ”clustering” est utilisé.

2.3 Segmentation des coupes histologiques

Dans le cas de traitement de coupe histologiques et immunohistologiques, les valeurs des labels Y_i , $i = 1..n_{\text{labels}}$ sont associées à ”os”, ”cellules”, ”plasma (vide)”, ”globules rouges” ...

Le but de la classification est d’estimer à quel label est associé un pixel de l’image, c’est-à-dire d’indiquer s’il se situe dans du tissu osseux, du plasma, une cellule, un globule rouge... Segmenter une image consiste à associer à chaque pixel de l’image un label.

La figure 2.3 montre un exemple de segmentation sur une coupe histologique marquée à l’hématoxiline et éosine (H & E).

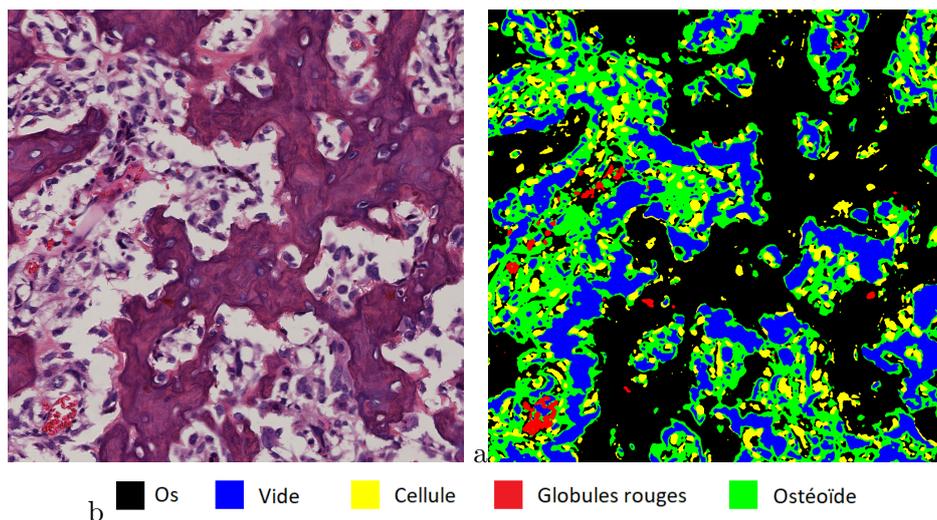


Figure 2.3: a) Partie d’une image de coupe histologique marquée H&E, b) classification par KNN, labels et couleurs utilisés pour l’illustration de la segmentation indiqués dans la légende.

Les images étudiées sont de très grande taille ($50000px \times 50000px$ en moyenne, taille caractéristique du pixel $0.466\mu m$) et nous recherchons un compromis entre la précision de l’algorithme et le temps de calcul. Dans un premier temps, nous allons réaliser une segmentation basée sur la couleur du pixel, avec des pré- et post-traitement. Des traitements supplémentaires

ultérieurement développés seront présentés au chapitre 3.

2.4 Classification en fonction de la couleur des pixels

Dans cette partie, les pixels seront classés par la seule donnée de leur couleur via une classification dans le cube RGB (Red r , Green g , Blue b). La couleur d'un pixel est définie par un triplet de trois entiers, illustrée figure 2.4, avec:

$$(r, g, b) \in \llbracket 0, 255 \rrbracket^3. \quad (2.2)$$

Les observations X_i seront alors les valeurs RGB de la couleur de chaque pixel.

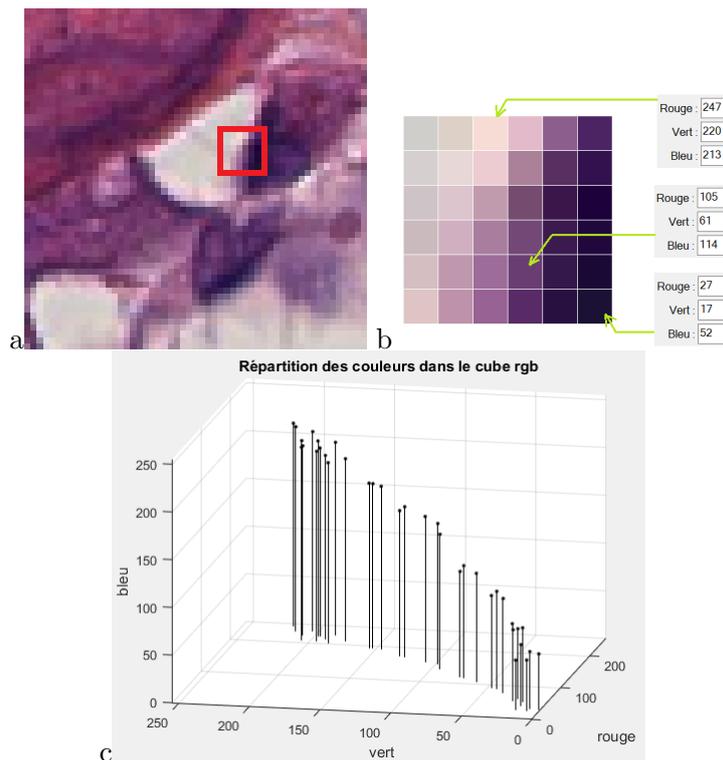


Figure 2.4: a) Partie d'une image histologique (taille caractéristique du pixel $0.466\mu m$, b) zoom sur quelques pixels de cette image, et c) répartition de la couleur de ces pixels dans le cube RGB.

Il y a un intérêt particulier à se baser uniquement sur la couleur pour limiter le temps de calcul au vu de la taille des images. Le fait de se restreindre à la couleur permet de s'appuyer sur un unique classement des éléments du cube RGB, soit $16,9 \times 10^6$ couleurs possibles. La segmentation de l'image consistera alors à associer le label du cube RGB à la couleur correspondante.

2.5 Visualisation de la répartition des couleurs des pixels

Afin d'évaluer la complexité du problème, il est possible de visualiser la répartition des couleurs des pixels d'une image dans le cube RGB. L'espace des couleurs du cube RGB étant en 3 dimensions, il ne se pose pas, pour l'instant, de question de réduction de dimensions.

La figure 2.5 montre la répartition des couleurs dans le cube RGB en ayant échantillonné à la main des zones d'images dont on sait déterminer la nature, c'est-à-dire, du tissu osseux, du plasma (vide), des cellules, des globules rouges ou des fibres roses (ground truth).

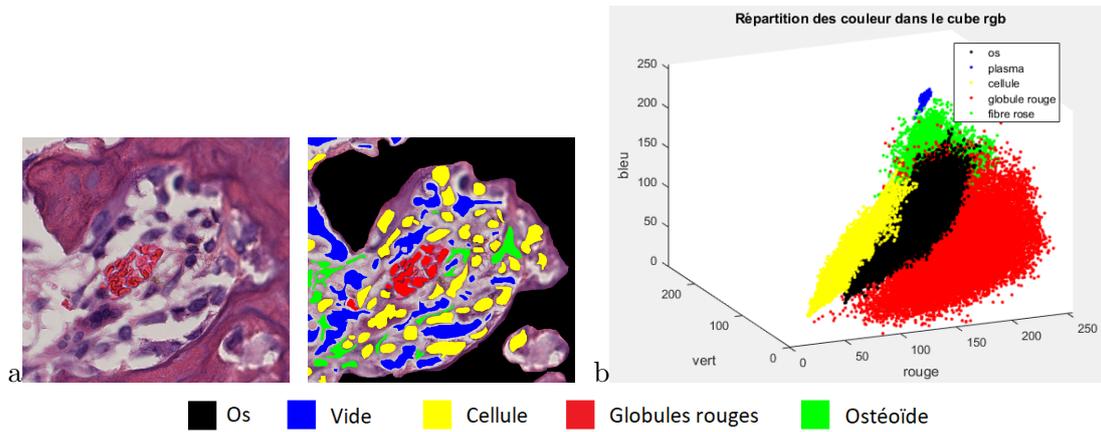


Figure 2.5: a) Echantillonnage manuel sur une coupe histologique, b) répartition des couleurs dans le cube RGB.

Chaque label occupe une zone spécifique du cube RGB. Toutefois, des chevauchements apparaissent, notamment dans les zones représentant les zones osseuses et les cellules. Nous discuterons ce point dans le chapitre suivant de ce document.

Dans la suite, la segmentation va consister à chercher des surfaces de séparation entre ces espaces labellisés du cube RGB par des méthodes de classification.

2.6 Quelques exemples de classification des images avec des algorithmes non-supervisés et supervisés

La méthode k-means

Le k-means est un algorithme de machine learning **non supervisé** et par conséquent ne requiert pas de données d'apprentissage et donc peut être utilisé sans l'intervention d'un observateur expert. L'algorithme du k-means cherche à minimiser la quantité:

$$\sum_{i=1}^k \sum_{x \in A_i} |x - c_i|^2 \quad (2.3)$$

où x est une observation regroupant des données et les c_i pour $i \in [1, k]$ sont k centres dans l'espace des observations et $(A_i)_i$ la partition de Voronoï associée à $(c_i)_i$ [57], avec:

$$A_i = \{x, \forall j \neq i, |x - c_i| \leq |x - c_j|\} \quad (2.4)$$

L'algorithme itératif est initié par k centres choisis aléatoirement parmi les observations. On détermine la partition de Voronoï associée à ces centres. Les centres de l'itération suivante sont les isobarycentres des ensembles de cette partition. Nous commençons par tester l'algorithme du k-means sur une image histologique de taille $7569px \times 8928px$, (figure 2.6).

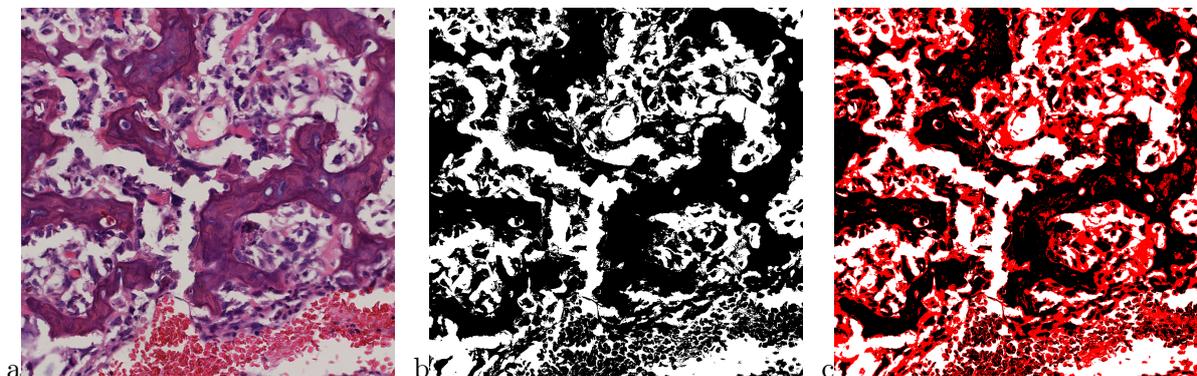


Figure 2.6: a) Zoom sur une coupe histologique H&E, b) résultat de la segmentation par k-means avec 2 centres, labels noir et blanc, c) résultat de la segmentation par k-means avec 3 centres, labels noir, blanc, rouge.

Les temps d'exécution d'apprentissage et de classification sont assez importants et croissent avec le nombre de centres. Il est possible de réduire la taille des données par tirage aléatoire. Cependant, compte tenu la répartition des couleurs dans le cube RGB, figure 2.5), appliquer l'algorithme du k-means ne permet d'obtenir une segmentation fine avec peu de centres, figure 2.6. Il est possible tester cet algorithme avec un nombre plus élevé de centres (figure 2.7) et de fusionner par la suite certains clusters.

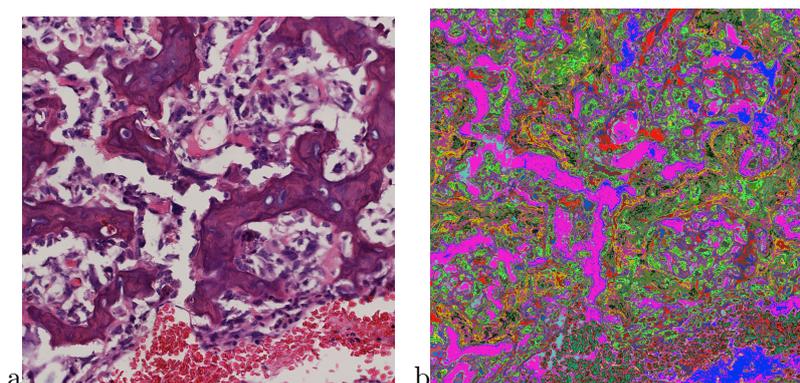


Figure 2.7: a) Zoom sur une coupe histologique H&E, b) clusters obtenus avec la méthode k-means et considérant 30 centres.

La figure 2.7 a été obtenue en utilisant une classification hiérarchique ascendante et 30 centres, des tables de contingence ont également été testées. Malgré certains progrès, les résultats de segmentation des images n'ont pas été satisfaisants. En effet, on n'observe pas de groupes isolés de manière générale dans la répartition des couleurs dans le cube RGB pour les coupes histologiques.

D'autres algorithmes non supervisés tels que le spectral clustering ont été testés, mais leur exécution demande un grand temps de calcul (de distance et de diagonalisation), et les résultats n'étaient également pas satisfaisants.

Méthode des Random Forest et AdaBoost

L'algorithme de Random Forest est un algorithme avec apprentissage **supervisé**, il donne un label sur une observation en effectuant un vote sur les labels renvoyés par un ensemble d'arbres de décision [57]. Il est largement utilisé dans les logiciels opensource car populaire du fait de sa robustesse [54]. L'AdaBoost est semblable au Random Forest, mais donne un poids plus élevé aux observations mal classées à l'étape précédente et réitère le Random Forest avec les poids mis à jour [57]. Nous avons testé ces algorithmes sur une coupe histologique de taille $7568px \times 8928px$.

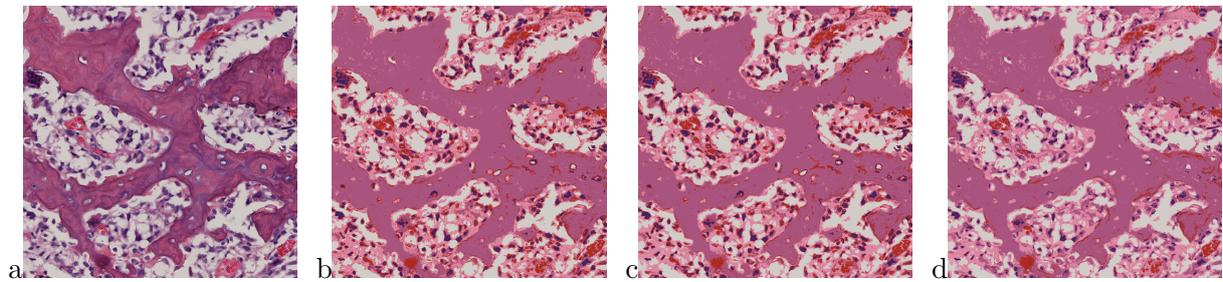


Figure 2.8: a) Image initiale, b) segmentation par random forest avec 10 arbres de décision, c) segmentation par random forest avec 100 arbres de décision et d) segmentation par Adaboost. L'os est segmenté en rose, les cellules en violet, les globules rouges en rouge, les fibres en rose claire et le milieu interstitiel en blanc.

Le résultat renvoyé par le Random Forest est assez bruité comme on peut le voir sur la figure 2.8. On peut émettre une hypothèse sur l'origine de ce bruit. Bien que le vote sur tous les arbres permette d'affiner la frontière dans le cube RGB entre les différents labels, cette frontière reste constituée par des morceaux de plans parallèles aux plans d'équation $x_k = c$, figure 2.9, ce qui n'est pas pertinent pour les configurations de chevauchement comme montré sur la figure 2.5.

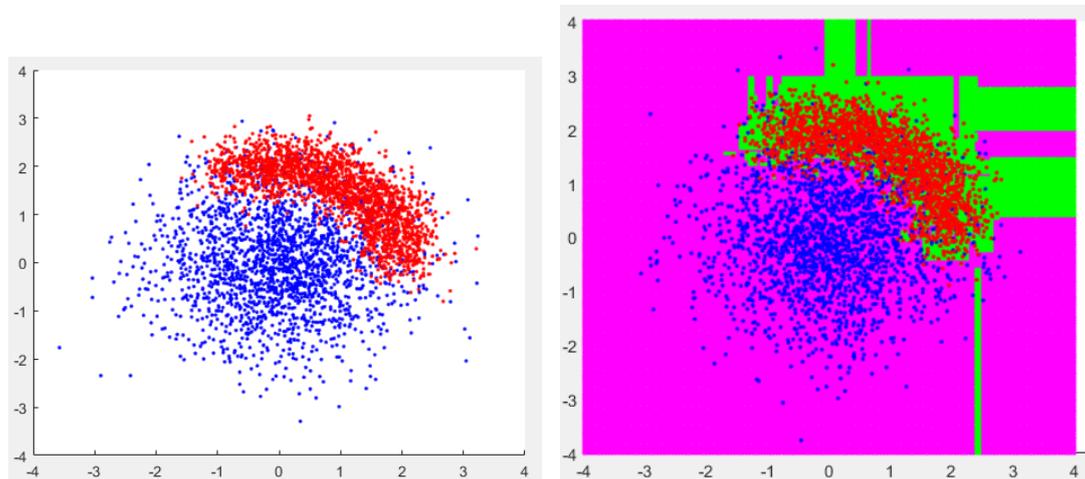


Figure 2.9: Exemple de classification résultant de l'algorithme de Random Forest, ici avec 10 arbres de décision.

Sur des zones atteintes par l'image mais non représentées par l'échantillon, la classification peut devenir très fluctuante. Or, les couleurs rencontrées dans la phase osseuse qui nous intéresse présente une grande variabilité en fonction de la position dans l'image. La figure 2.10 montre des variations de couleurs issues d'une même image.

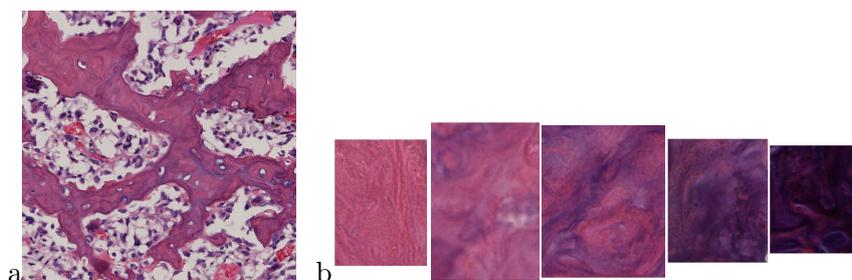


Figure 2.10: a) Image initiale, b) variations de teintes de la phase osseuse dans une même image.

Nous voyons donc qu'il n'est pas possible de faire un échantillonnage exhaustif. De plus, étant donné l'influence de la taille des données d'entraînement sur le temps d'apprentissage, il

est préférable de choisir un algorithme qui limite la variance sur les zones non explorées.

Méthode des K plus proches voisins (KNN)

Le KNN est un algorithme de classification **supervisée** pour lequel une nouvelle observation est classée en fonction des couleurs des k plus proches voisins et le label donné à cette observation est le label majoritaire de ces k voisins [57]. Il faut donc effectuer un échantillonnage qui constitue les données d'entraînement (petit échantillon des différentes phases que l'on cherche à classer). Cet algorithme est une version améliorée d'une partition de Voronoï, qui peut être retrouvée lorsque $k = 1$. Le fait d'avoir k voisins diminue la variance qui pourrait être obtenue avec des méthodes plus simples. Une illustration de la segmentation par la méthode KNN est donnée figure 2.3.

Selon les parties de l'image, le KNN arrive à distinguer l'os des cellules (figure 2.11).

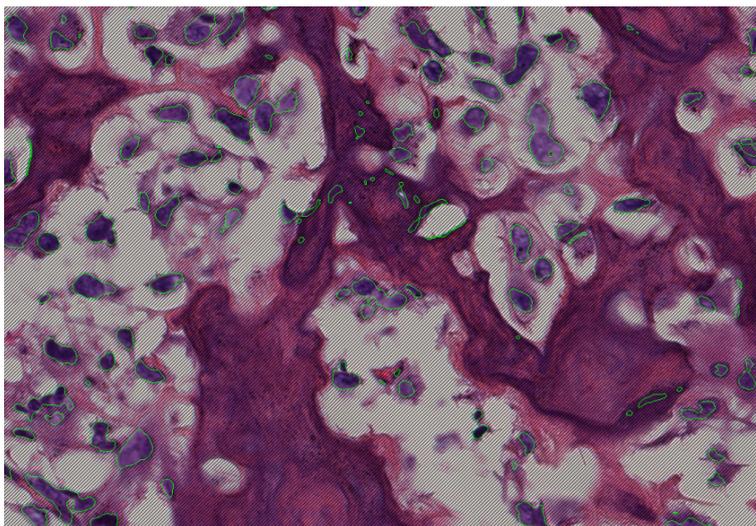


Figure 2.11: Frontière des cellules en vert obtenue par méthode KNN et partie hors cellules hachurée.

Comparaison des algorithmes

Les algorithmes de classification supervisés sont testés sur des images immunohistologiques qui sont plus contraignantes, comme illustré sur la figure 2.12.

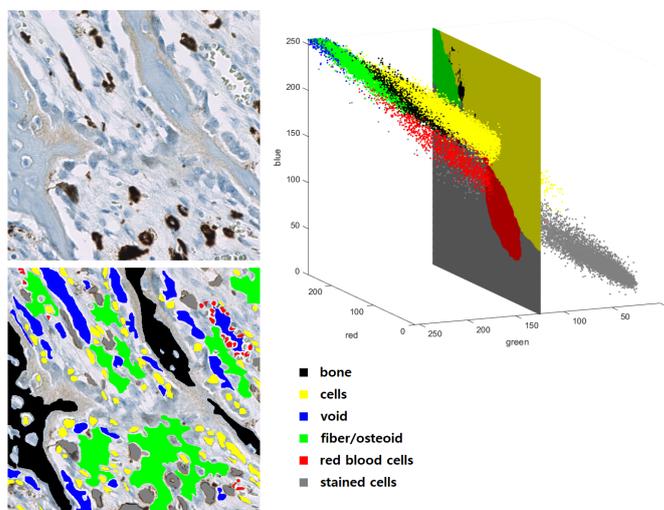


Figure 2.12: L'image du haut correspond à un zoom sur une section d'OST colorée avec CD68, l'image du bas montre une segmentation manuelle de la section par un expert avec les labels associés. Le nuage de points montre la distribution 3D dans le cube RGB des labels obtenus par segmentation manuelle.

Les pourcentages indiqués dans les courbes 2.13 et 2.14 ainsi que dans le tableau de conclusion du chapitre correspondent à un indice appelé Intersection-Over-Union (IoU, Jaccard Index) très similaire et souvent égal au Dice score. Ces indices, utilisés pour évaluer la qualité de la segmentation, caractérisent le pourcentage de surfaces correspondant à l'intersection de la ground truth ou segmentation manuelle effectuée par l'observateur expert et le résultat de la segmentation numérique.

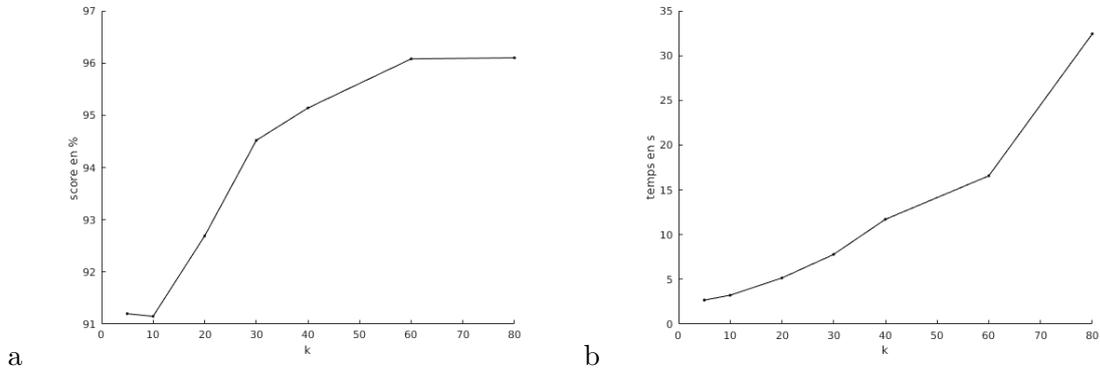


Figure 2.13: KNN, **a)** score en % en fonction de k , **b)** temps d'exécution en secondes en fonction de k .

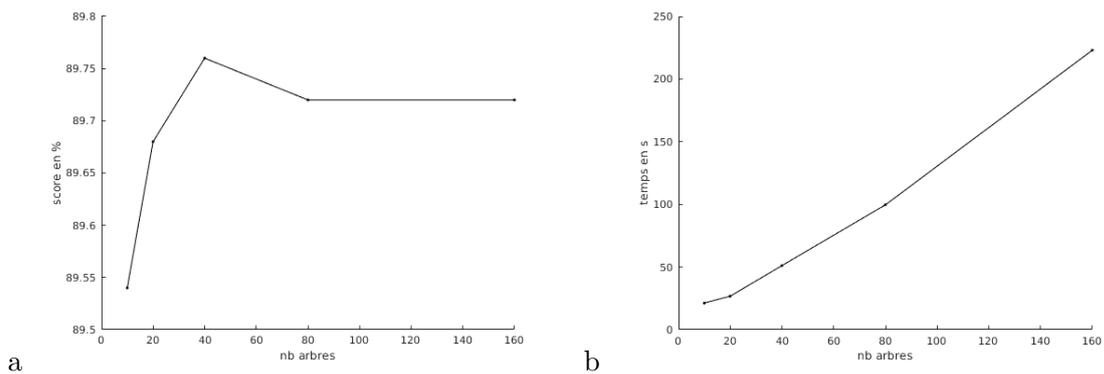


Figure 2.14: Random Forest, **a)** score en % en fonction du nombre d'arbres, **b)** temps d'exécution en secondes en fonction du nombre d'arbres.

La comparaison des figures 2.13 et 2.14 a permis de sélectionner la méthode KNN comme méthode de segmentation privilégiée. Notons que quelques tests ont été réalisés avec des méthodes de deep learning [57]¹ mais cette stratégie a été rapidement abandonnée au vu de la variabilité inter et intra-tumorale des images. En effet, le nombre de données d'entraînement nécessaires au deep learning ne pouvait être atteint dans le contexte de ce travail de thèse.

2.7 Segmentation utilisant l'algorithme KNN

En lien avec la méthode KNN, cette partie décrit la forme de l'échantillonnage (extraction des données d'entraînement), les stratégies de réduction du temps d'apprentissage et d'obtention d'un échantillonnage pertinent.

Un entraînement s'appuie sur un échantillonnage. L'échantillon est la donnée de deux vecteurs X et Y définis ci-dessous avec n le nombre d'observations dans l'échantillon. Le vecteur X contient les observations, dans notre cas ce sont des couleurs. Soit:

¹Apprentissage automatique tentant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} r_1 & g_1 & b_1 \\ r_2 & g_2 & b_2 \\ \vdots & \vdots & \vdots \\ r_n & g_n & b_n \end{pmatrix} \quad (2.5)$$

$$X_i = (r_i, g_i, b_i) \in [[0, 255]]^3 \quad (2.6)$$

où r_i , g_i et b_i sont les composantes rouges, vertes et bleues du i ème pixel de l'échantillon. Le vecteur Y contient les labels associés aux observations, où n_{labels} est le nombre de labels différents :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n_{\text{labels}}} \end{pmatrix} \quad (2.7)$$

$$Y_i \in [[1, n_{\text{labels}}]] \quad (2.8)$$

Nous proposons ici une solution pour réduire la taille des données à classer dans le cube RGB, le temps d'apprentissage du KNN étant linéaire avec la taille des données.

Soit \tilde{X} le vecteur parcourant les couleurs du cube RGB et N le nombre de couleurs à classer :

$$\tilde{X} = \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_N \end{pmatrix} = \begin{pmatrix} \tilde{r}_1 & \tilde{g}_1 & \tilde{b}_1 \\ \tilde{r}_2 & \tilde{g}_2 & \tilde{b}_2 \\ \vdots & \vdots & \vdots \\ \tilde{r}_N & \tilde{g}_N & \tilde{b}_N \end{pmatrix} \quad (2.9)$$

La taille du vecteur \tilde{X} est de $256^3 \times 3$. Cependant il n'est pas pertinent de classer toutes ces couleurs, car certaines ne sont jamais présentes dans l'image à segmenter, comme illustré figure 2.5.

Une méthode pour réduire le nombre N de couleurs à classer est de diviser le cube RGB en sous-espaces de petite dimension, par exemple des cubes de côtés 16. Un parcours de l'image de la coupe histologique permet de savoir quels sont les cubes vraiment atteints par l'image. Il suffit ensuite de conserver un vecteur de booléens donnant la présence ou non des 4096 petits cubes pour reconstituer la classification dans le cube RGB. Le temps d'entraînement et la taille de certains tableaux seront donc réduits de manière importante du fait de la réduction de la taille des données à classer.

L'entraînement du KNN prend en argument les vecteurs X et \tilde{X} , correspondant respectivement aux données d'entraînement et aux données à classer, et le nombre de voisins k . Les informations renvoyées par le KNN sont deux matrices. Une matrice M_{indice} de taille $N \times k$ contenant les indices des k plus proches voisins parmi l'échantillon pour chaque couleur à classer, soit:

$$M_{\text{indice}} = \begin{pmatrix} i_{1,1} & \cdots & i_{1,k} \\ i_{2,1} & \cdots & i_{2,k} \\ \vdots & & \vdots \\ i_{N,1} & \cdots & i_{N,k} \end{pmatrix} \quad (2.10)$$

Ainsi, pour la j -ième couleur à classer \tilde{X}_j , les k plus proches voisins parmi l'échantillon sont $X_{i_{j,1}}$, $X_{i_{j,2}}$, ... et $X_{i_{j,k}}$. Ces k voisins sont rangés du plus proche au plus éloigné. La deuxième matrice M_{distance} , elle aussi de taille $N \times k$, contient les k distances des k plus proches voisins pour chaque couleur à classer :

$$M_{\text{distance}} = \begin{pmatrix} d_{1,1} & \cdots & d_{1,k} \\ d_{2,1} & \cdots & d_{2,k} \\ \vdots & & \vdots \\ d_{N,1} & \cdots & d_{N,k} \end{pmatrix} \quad (2.11)$$

Enfin, si on veut obtenir la classification sur les données \tilde{X} , il faut créer la matrice M_{labels} , soit:

$$M_{\text{labels}} = \begin{pmatrix} Y_{i_{1,1}} & \cdots & Y_{i_{1,k}} \\ Y_{i_{2,1}} & \cdots & Y_{i_{2,k}} \\ \vdots & & \vdots \\ Y_{i_{N,1}} & \cdots & Y_{i_{N,k}} \end{pmatrix} \quad (2.12)$$

Le vecteur \tilde{Y} qui classe les données \tilde{X} en déterminant le label majoritaire sur chaque ligne de la matrice M_{labels} est:

$$\tilde{Y} = \begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \vdots \\ \tilde{Y}_N \end{pmatrix} \quad (2.13)$$

Ainsi, la j -ième couleur \tilde{X}_j sera classée avec le label \tilde{Y}_j . Il est possible d'établir une fonction qui donne directement le vecteur \tilde{Y} , mais les matrices M_{indice} , M_{distance} et M_{labels} vont être utiles dans la suite des développements.

Un algorithme "naïf" du KNN ferait intervenir un grand nombre de calculs de distances ce qui serait coûteux numériquement. Cependant, il existe des algorithmes d'optimisation du KNN qui séparent les données d'entraînement en clusters ce qui limite les calculs de distance (inclus dans la bibliothèque de machine learning de Matlab©).

Par ailleurs, il est possible de rajouter des données d'entraînement au KNN sans qu'il soit nécessaire de reprendre l'apprentissage initial. Supposons que nous ayons déjà utilisé le KNN pour un échantillonnage X^1 et que nous voudrions ajouter de nouvelles données X^2 dans l'échantillon.

Soient n_1 et n_2 le nombre d'observations des échantillons X^1 et X^2 et soit X la concaténation de ces deux vecteurs:

$$X = \begin{pmatrix} X^1 \\ X^2 \end{pmatrix} \quad (2.14)$$

On souhaite obtenir la classification avec l'échantillon total X .

Il est possible de faire un nouvel apprentissage sur les $n_1 + n_2$ données de X , ce qui implique de traiter à nouveau les n_1 données comprises dans X^1 . Mais, il est également possible de s'affranchir de X^1 , ce qui peut être un important gain de temps de calcul selon les valeurs de n_1 et n_2 .

En effet, nous définissons quatre matrices M_{indice}^1 , M_{distance}^1 , M_{indice}^2 et M_{distance}^2 de taille $N \times k$:

$$M_{\text{indice}}^1 = \begin{pmatrix} i_{1,1}^1 & \cdots & i_{1,k}^1 \\ i_{2,1}^1 & \cdots & i_{2,k}^1 \\ \vdots & & \vdots \\ i_{N,1}^1 & \cdots & i_{N,k}^1 \end{pmatrix} \quad M_{\text{indice}}^2 = \begin{pmatrix} i_{1,1}^2 & \cdots & i_{1,k}^2 \\ i_{2,1}^2 & \cdots & i_{2,k}^2 \\ \vdots & & \vdots \\ i_{N,1}^2 & \cdots & i_{N,k}^2 \end{pmatrix} \quad (2.15)$$

$$M_{\text{distance}}^1 = \begin{pmatrix} d_{1,1}^1 & \cdots & d_{1,k}^1 \\ d_{2,1}^1 & \cdots & d_{2,k}^1 \\ \vdots & & \vdots \\ d_{N,1}^1 & \cdots & d_{N,k}^1 \end{pmatrix} \quad M_{\text{distance}}^2 = \begin{pmatrix} d_{1,1}^2 & \cdots & d_{1,k}^2 \\ d_{2,1}^2 & \cdots & d_{2,k}^2 \\ \vdots & & \vdots \\ d_{N,1}^2 & \cdots & d_{N,k}^2 \end{pmatrix} \quad (2.16)$$

Le but est de déterminer les matrices M_{indice} et M_{distance} à partir de ces 4 matrices permettant de classer l'échantillon X . Remarquons que l'indice i associé à X^2 deviendra l'indice $i + n_1$ pour le vecteur X .

Une première possibilité est de concaténer les matrices M_{indice}^1 et M_{indice}^2 , et, M_{distance}^1 et M_{distance}^2 :

$$\tilde{M}_{\text{indice}} = \left(M_{\text{indice}}^1 \quad (M_{\text{indice}}^2 + n_1) \right) = \begin{pmatrix} i_{1,1}^1 & \cdots & i_{1,k}^1 & i_{1,1}^2 + n_1 & \cdots & i_{1,k}^2 + n_1 \\ i_{2,1}^1 & \cdots & i_{2,k}^1 & i_{2,1}^2 + n_1 & \cdots & i_{2,k}^2 + n_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ i_{N,1}^1 & \cdots & i_{N,k}^1 & i_{N,1}^2 + n_1 & \cdots & i_{N,k}^2 + n_1 \end{pmatrix} \quad (2.17)$$

$$\tilde{M}_{\text{distance}} = \left(M_{\text{distance}}^1 \quad M_{\text{distance}}^2 \right) = \begin{pmatrix} d_{1,1}^1 & \cdots & d_{1,k}^1 & d_{1,1}^2 & \cdots & d_{1,k}^2 \\ d_{2,1}^1 & \cdots & d_{2,k}^1 & d_{2,1}^2 & \cdots & d_{2,k}^2 \\ \vdots & & \vdots & \vdots & & \vdots \\ d_{N,1}^1 & \cdots & d_{N,k}^1 & d_{N,1}^2 & \cdots & d_{N,k}^2 \end{pmatrix} \quad (2.18)$$

Ceci est suivi du tri des $2k$ distances $(d_{j,1}^1 \cdots d_{j,k}^1 \quad d_{j,1}^2 \cdots d_{j,k}^2)$ de chaque ligne de $\tilde{M}_{\text{distance}}$, puis d'une permutation sur les indices. On ne garde alors que les k premières valeurs de chaque ligne. Cette approche est coûteuse.

Il est important de noter que les lignes de M_{distance}^1 et M_{distance}^2 sont déjà triées. Il est donc possible d'appliquer la même méthode utilisée lors de la remontée de l'algorithme de tri-fusion illustré ci-dessous sur deux vecteurs ligne de dimension 4, triés. De plus, il est possible de ne trier que les k premiers termes.

Pour rappel, une des étapes de l'algorithme de tri-fusion visant à trier un tableau en des temps de l'ordre de la *dimension * log(dimension)* est indiquée ci-dessous.

Supposons que l'on ait deux vecteurs ligne L_1 et L_2 :

$$L_1 = (1 \ 4 \ 5 \ 9) \quad L_2 = (2 \ 3 \ 7 \ 8) \quad (2.19)$$

On souhaite ranger les quatre premiers éléments dans un vecteur L . On compare d'abord les deux premiers éléments de L_1 et L_2

$$L_1 = (\underset{\uparrow}{1} \ 4 \ 5 \ 9) \quad L_2 = (2 \ \underset{\uparrow}{3} \ 7 \ 8) \quad (2.20)$$

Le plus petit nombre est 1, on range 1 dans L , puis on avance dans le vecteur L_1 .

$$L = (1 \ \dots \ \dots \ \dots) \quad (2.21)$$

$$L_1 = (1 \ \underset{\uparrow}{4} \ 5 \ 9) \quad L_2 = (2 \ \underset{\uparrow}{3} \ 7 \ 8) \quad (2.22)$$

Le plus petit nombre est 2.

$$L = (1 \ 2 \ \dots \ \dots) \quad (2.23)$$

$$L_1 = (1 \ \underset{\uparrow}{4} \ 5 \ 9) \quad L_2 = (2 \ \underset{\uparrow}{3} \ 7 \ 8) \quad (2.24)$$

Ainsi de suite,

$$L = (1 \ 2 \ 3 \ \dots) \quad (2.25)$$

$$L_1 = (1 \underset{\uparrow}{4} 5 9) \quad L_2 = (2 3 \underset{\uparrow}{7} 8) \quad (2.26)$$

$$L = (1 2 3 4) \quad (2.27)$$

Il est donc possible de faire ce tri simultanément sur toutes les lignes de M_{distance}^1 et M_{distance}^2 de manière matricielle.

Prenons l'exemple d'un langage de programmation dont :

- l'indexation des tableaux commence à 1
- une matrice (de taille $N \times k$) peut être lue colonne par colonne lorsqu'on la décrit avec un seul indice:

$$M_{i,j} = M(i + N(j - 1)) \quad (2.28)$$

On obtient un vecteur colonne si i et j sont remplacés par des vecteurs.

- $M(:, j)$ désigne la j -ième colonne de M
- Soient v et w deux vecteurs colonne de même taille, $b < a$ renvoi le vecteur booléen $(b_i < a_i)_i$.
- Sommer et multiplier terme à terme des tableaux de nombres et de booléens avec la valeur "true" comptant pour 1 et "false" pour 0.
- $1 : N$ est le vecteur colonne allant de 1 jusqu'à N .

La méthodologie peut être résumée ci-dessous:

Initialisation :

$$v^1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad v^2 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad (2.29)$$

v^1 et v^2 sont des vecteurs de taille N , ils vont contenir les indices analogues aux "flèches" utilisées dans l'exemple avec L_1 et L_2 .

Pour j allant de 1 à k , on définit a un vecteur des distances des matrices M aux indices v et b un vecteur booléen (avec 1 pour vrai et 0 pour faux) tels que:

$$a^1 = M_{\text{distance}}^1(1 : N + N(v^1 - 1)) \quad a^2 = M_{\text{distance}}^2(1 : N + N(v^2 - 1)) \quad (2.30)$$

$$b = a^1 < a^2 \quad (2.31)$$

$$M_{\text{distance}}(:, j) = a^1 .* b + a^2 .* (1 - b) \quad (2.32)$$

$$M_{\text{indice}}(:, j) = M_{\text{indice}}^1(1 : N + N(v^1 - 1)).*b + M_{\text{indice}}^2(1 : N + N(v^2 - 1)).*(1 - b) \quad (2.33)$$

$$v^1 = v^1 + b \quad v^2 = v^2 + (1 - b) \quad (2.34)$$

Ce processus doit être itéré jusqu'à ce que $j = k$, le nombre de voisins choisis pour la méthode KNN.

Ainsi l'ajout d'échantillons ne demande pas de refaire tous les calculs depuis le début. En effet, en réorganisant par une étape de la fusion de la méthode de tri-fusion combinant les matrices de distance initiales et du nouvel échantillonnage, la matrice de distance totale est obtenue très rapidement. A notre connaissance, cette approche n'a pas été discutée à ce jour dans la littérature et constitue une des contributions de cette thèse au domaine du traitement d'images.

Par ailleurs, un échantillonnage peut être itératif tel que développé ci-dessous. Lorsqu'on applique le KNN, on est tenté de fournir une taille conséquente de données d'entraînement pour obtenir la meilleure segmentation possible. Cependant, le temps d'exécution de l'apprentissage croît avec la taille des données d'apprentissage.

La possibilité d'ajouter au fur et à mesure de nouvelles données d'entraînement nous permet d'améliorer la qualité de la segmentation de manière itérative. Le but n'est pas seulement de faire un simple "essais/erreur" sur l'échantillonnage, mais de faire des ajouts d'échantillons de

manière pertinente et d'éviter certaines redondances. Ceci est possible par l'utilisation des quantités guidant le choix des échantillons additionnels spécifiques.

Premièrement, la matrice M_{distance} donne les k distances entre la couleur du cube RGB classée et les k voisins dans les données d'entraînement. Il est possible de parcourir les couleurs de l'image à segmenter et de leur associer la première colonne de la matrice M_{distance} .

Dans l'exemple figure 2.15, nous omettons l'échantillonnage des cellules marquées en marron dans une image d'immunohistologie, alors la carte des distances permet d'identifier les zones de l'image qui sont mal représentées dans l'échantillon et de rajouter les données d'entraînement pertinentes dans l'échantillonnage (échantillons marron dans figures du bas). On voit alors que la carte des distances donne des amplitudes proches de zéro indiquant que l'algorithme est capable d'associer chaque zone de l'image à un label.

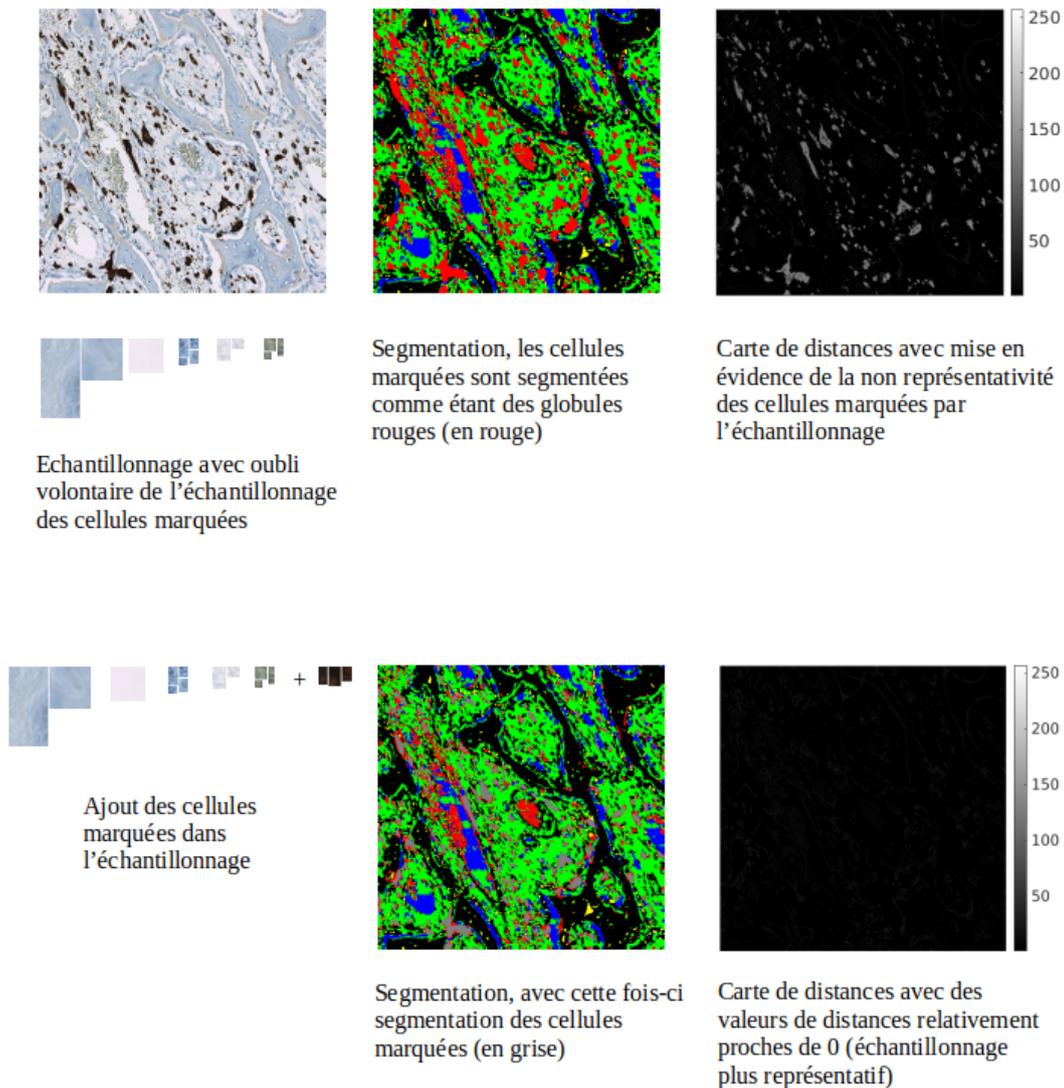


Figure 2.15: Images haut: coupe histologique originale, segmentation par KNN obtenue avec les échantillons indiqués en dessous de la coupe, carte des distances obtenue en omettant d'entraîner les cellules marquées en marron. Images bas: segmentation par KNN obtenue en rajoutant l'entraînement sur les cellules marron, carte des distances obtenue avec ce nouvel échantillon.

Étant donné que la classification ne s'appuie que sur la couleur, les pixels sont classés individuellement. Des stratégies de post-traitement dites "simples" peuvent être appliquées pour harmoniser cette segmentation afin d'obtenir des zones segmentées moins bruitées. Un premier post-traitement présenté est la suppression de partie connexes utile, par exemple lorsque l'on

cherche à extraire des métriques portant sur la phase osseuse. Le second post-traitement est le lissage de frontières entre labels utilisé pour améliorer les scores de la segmentation sans temps de calcul supplémentaire.

Après la segmentation par la méthode KNN, un post-traitement peut être appliqué consistant à supprimer les pixels formant de petites parties connexes. Cet algorithme est basé sur un seuillage de taille des parties connexes. On applique ce traitement pour les parties connexes de type tissu osseux puis à nouveau pour les parties connexes des autres phases pour obtenir une phase osseuse peu bruitée (figure 2.16).

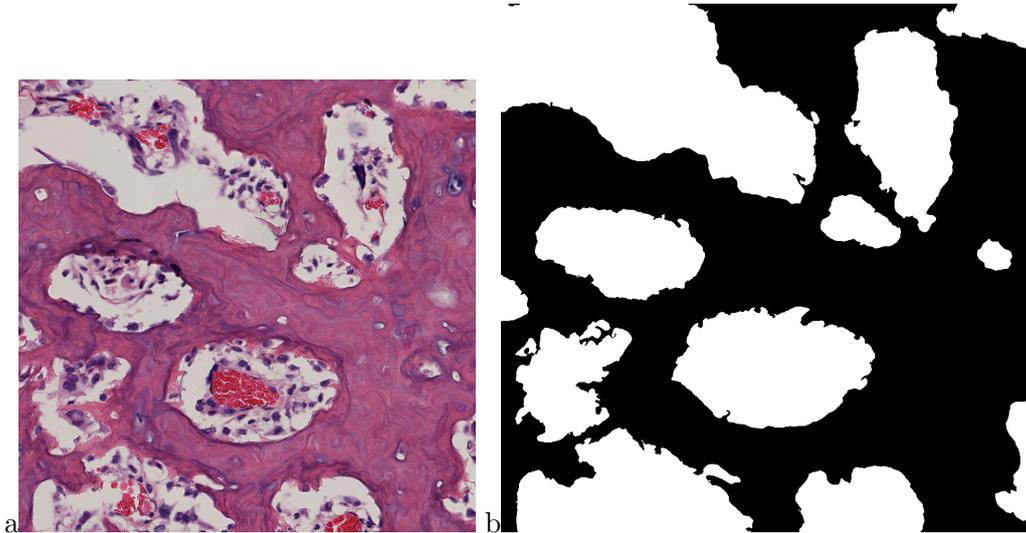


Figure 2.16: a) Image initiale, b) segmentation KNN, binarisation portant sur la phase osseuse et suppression des parties connexes.

Après classification des pixels par la couleur, certains pixels isolés peuvent être mal classés. Il est alors possible de lisser la segmentation en remplaçant le label d'un pixel par le label le plus représenté sur un voisinage de ce pixel pour un coût numérique quasi-nul. L'influence de ce lissage est illustré figure 2.17 pour une approche KNN. On peut observer des optimaux de score pour des largeurs de voisinage dans le lissage entre 2 et 4, pour des k variant ici de 5 à 80.

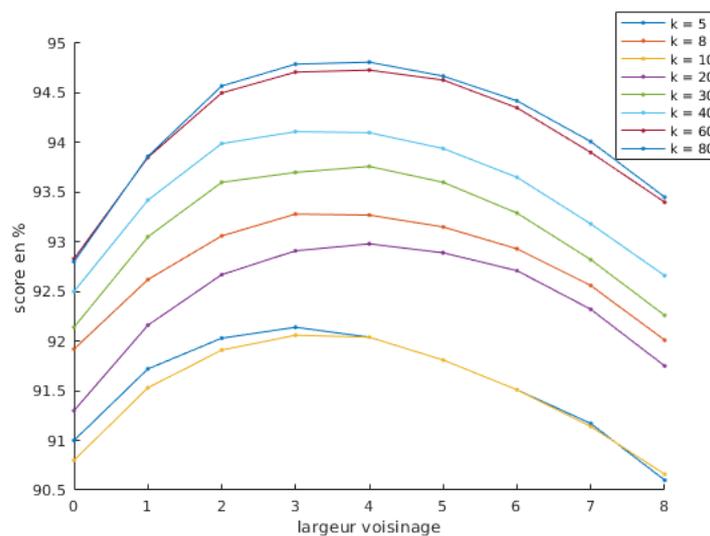
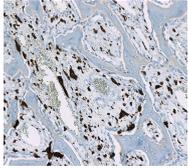
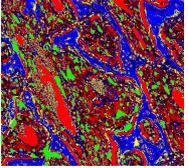
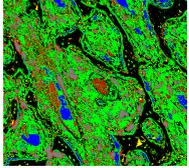
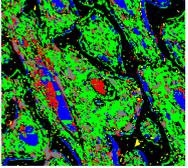


Figure 2.17: Score en % en fonction de la largeur du voisinage pour le lissage, courbes avec différents k pour la méthode KNN.

2.8 Conclusion

Nous avons présenté dans ce chapitre, certains aspects principaux associés à la segmentation de coupes histologiques et des développements adaptés à l'étude de l'OST. Pour présenter une synthèse explicite, nous avons choisi les résultats issus de cas tests relativement génériques. D'autres cas ont été explorés et cela a abouti à des conclusions similaires. Une coupe marquée CD68 est étudiée. Le tableau 2.1 présente les scores et le temps de calcul obtenus pour une segmentation par méthode K-means non supervisée, Random Forest et KNN supervisées. Il apparaît de manière explicite, la supériorité de l'algorithme KNN donnant le meilleur compromis temps de calcul et score (temps CPU inférieur à 1h attendu). L'algorithme KNN sera sélectionné pour la suite des travaux avec un nombre de voisins $k = 10$.

	K-means (unsupervised)	Random Forest (supervised)	KNN (supervised)
			
Parameters	6 clusters	20 trees	k=15
CPU time	15.30 s	28.23 s	5.70 s
Score test	(*)	89.77 %	91.55 %
Score + lissage largeur 4		91.15 %	93.04 %
lisse + taille parties connexes (s=30000)			95.23 %

(*) :

labels and %	1	2	3	4	5	6
os	0	10.2	0	81.96	7.82	0.02
vide	100	0	0	0	0	0
cellules	0	74.91	0	10.42	0.32	14.35
globules rouges	0.15	32.94	0	2.56	0.29	64.06
fibres	49.36	0	0	2.97	47.67	0
cellules marquées	0	0	99.61	0	0	0.39

Table 2.1: Comparaison des approches k-means, random forest et KNN pour une coupe immunohistologique (marquage CD68). Un tableau séparé est indiqué pour les scores du K-means indiquant le pourcentage des surfaces de chaque phase associé aux clusters de 1 à 6. Seules les cellules marquées peuvent être correctement identifiées par cette méthode.

Chapter 3

Recherche de filtres pour l'amélioration de la segmentation

Contents

3.1	Nomenclature	30
3.2	Introduction : filtres linéaires	31
3.2.1	Exemple de filtres linéaires simples approchant des dérivées partielles	31
3.2.2	Définition	32
3.2.3	Compréhension des filtres linéaires: exemple des filtres issus de la SVM	34
3.3	Décomposition d'un filtre linéaire sur un signal "lisse"	36
3.3.1	Approximation de la dérivée partielle ∂^γ	36
3.3.2	Exemple de l'approximation des filtres et comparaison avec les différences finies dans le cas 1D	39
3.3.3	Décomposition d'un filtre	41
3.4	Décomposition d'un filtre linéaire sur un signal bruité	43
3.4.1	Approximation de la dérivée partielle ∂^γ	43
3.4.2	Décomposition d'un filtre sur un signal bruité	46
3.5	Opérateurs différentiels invariants par rotation et symétrie	47
3.5.1	Introduction	47
3.5.2	Rotation de f	47
3.5.3	Opérateur différentiel φ	49
3.5.4	Recherche des opérateurs différentiels invariants par rotation et linéaires ($n=1$)	50
3.5.5	Recherche des opérateurs différentiels invariants par rotation quadratiques et homogènes ($n = 2, k_1 = k_2$)	52
3.5.6	Recherche des opérateurs quadratiques invariants par rotation et non homogènes ($n = 2, k_1 \neq k_2$)	54
3.5.7	Recherche des opérateurs différentiels invariants par rotation, de degré n et homogènes	56
3.5.8	Un exemple en degré 3 non homogène	58
3.5.9	Opérateurs invariants sur deux fonctions	59
3.6	Variance locale	60
3.6.1	Variance sur des fenêtres de l'image	60
3.6.2	Variance locale lisse sur l'image	61
3.7	Conclusion	62

Dans le chapitre 2, nous avons présenté une méthode de machine learning, la méthode KNN, qui donne des score de qualité (indice de Jaccard) à 95% sur les coupes HE et environ 90% sur les coupes immuno-histologiques. Malgré ces résultats encourageants, la distinction entre phase osseuse et cellulaire reste difficile sur certaines parties des images, en particulier immunohistologiques où les cellules superposées à la matrice osseuse ne sont pas identifiables par unique

approche KNN. Une classification, pour la segmentation, ne s'appuyant que sur la couleur d'un pixel ayant ses limites, nous allons proposer dans ce chapitre des améliorations en s'appuyant sur les données du voisinage des pixels à classer. Dans ce contexte, certaines quantités comme le gradient ou le laplacien peuvent ainsi donner des informations supplémentaires utiles pour l'amélioration de la segmentation. Une première idée est donc de superposer l'application de filtres linéaires à la méthode KNN.

Un filtre linéaire permet d'obtenir une nouvelle image à partir de l'image originale, mais également d'obtenir des informations complémentaires qui peuvent être utilisées pour différentes applications incluant la détection de contours ou la segmentation [61].

Dans ce chapitre, après avoir présenté quelques exemples de filtres linéaires, je montrerai comment obtenir un filtre approchant un filtre de dérivée partielle donnée, comment décomposer un filtre quelconque en combinaison linéaire d'approximation de dérivées partielles, comment prendre en compte le bruit et comment utiliser ces décompositions pour interpréter les filtres issus de la SVM (Support Vector Machine). Cette approche par décomposition de filtres par dérivées partielles offre deux avantages: l'interprétabilité des filtres par exemple issus de la SVM et la réduction de la dimension des données d'entraînement évitant la "malédiction de la dimensionnalité", réduisant le temps de calculs et permettant d'explorer également des modèles plus complexes comme la SVM non-linéaire, une approche très coûteuse en temps de calculs. De plus ces travaux ayant une visée translationnelle soumise aux réglementations des agences médicales, cette question de l'interprétabilité est cruciale. La deuxième partie de chapitre s'intéressera à la recherche de filtres invariants par rotation ou symétrie.

3.1 Nomenclature

$I = (i, j)$ coordonnées discrètes sur une image.

$X = (x_1, x_2)$ coordonnées continues en 2D.

u image lisse : application dans une domaine de \mathbb{R}^2 dans \mathbb{R} (monochrome) ou \mathbb{R}^3 (en couleur). $u(X)$ valeur de la couleur de u en $X = (x_1, x_2)$. Cette image n'existe pas concrètement.

$(u_I)_I$ image discrète (avec des pixels) : application de $[[1, N]] \times [[1, M]]$ dans \mathbb{R} (monochrome) ou \mathbb{R}^3 (en couleur). $u_I, u_{i,j}$ valeur de la couleur en $I = (i, j)$.

$(v_I)_I$ image bruitée.

h filtre sur une image discrète.

e_i vecteur unitaire valant 1 sur la i -ième coordonnée.

V voisinage sur une image (le plus souvent $[[-n, n]]^2$).

k ($k = (k_1, k_2)$) coordonnées dans le voisinage d'un pixel et dans le filtre (aussi noté m quand il y a besoin d'un deuxième "k" dans les sommes).

W ensemble des bi-indices des ordres de dérivées (le plus souvent $[[0, 2n]]^2$).

β, γ bi-indices des ordres des dérivées (aussi noté η quand il y a besoin d'un troisième bi-indice dans les sommes). Par exemple $\beta = (2, 1)$ correspond à la dérivée $\partial_x^2 \partial_y^1$.

$(\alpha_k)_{k \in V}$ filtre de coefficients α_k .

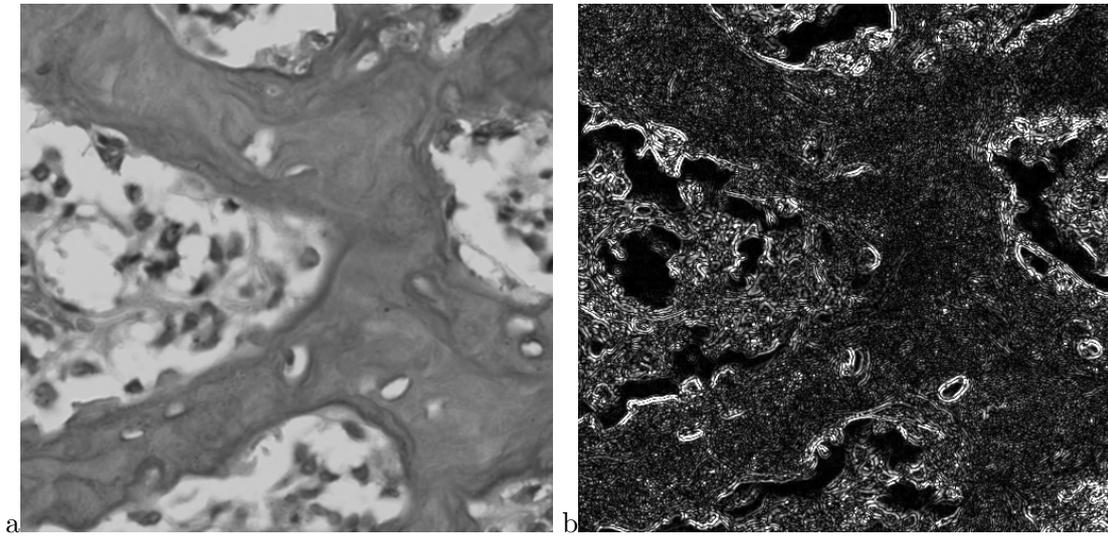


Figure 3.1: a) image initiale en niveaux de gris, b) image obtenue après application du filtre laplacien.

α_k^γ coefficient du filtre approximant la dérivée d'ordre γ en k .

ϕ erreur quadratique.

∂_X^γ notation multi-indice de la dérivée : $\partial_x^{\beta_1} \partial_y^{\beta_2}$.

φ opérateur, linéaire ou non, sur une image faisant intervenir des dérivées partielles.

3.2 Introduction : filtres linéaires

3.2.1 Exemple de filtres linéaires simples approchant des dérivées partielles

Nous définissons une image monochrome comme une application de $[[1, N]] \times [[1, M]]$ dans \mathbb{R} . N et M correspondent aux dimensions de l'image (largeur et hauteur) en nombre de pixels. Nous pouvons nommer $u_{i,j}$ la valeur de la couleur de l'image sur le pixel (i, j) .

Lorsque nous calculons le laplacien d'une image monochrome en utilisant la discrétisation du laplacien classique par différences finies, nous obtenons :

$$(\Delta u)_{i,j} \approx -4u_{i,j} + u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1} \quad (3.1)$$

une combinaison linéaire de $u_{i,j}$ et de certains de ses voisins.

En 2D, un filtre peut être représenté par une matrice de taille $p \times l$. Par exemple, l'approximation en différences finies du laplacien peut être représenté par la matrice 3×3 suivante:

$$h = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad (3.2)$$

Ce filtre linéaire est appliqué en chaque pixel (i, j) de l'image et donne une nouvelle image (figure 3.1).

Dans le traitement d'image, les filtres linéaires sont utilisés fréquemment, par exemple, pour les problèmes de détections de contours ou pour la segmentation.

Selon le filtre, on peut y reconnaître l'intervention de la valeur ou des dérivées de l'image. Par exemple :

- un filtre (1) de lissage classique (moyennage) a pour matrice :

$$\frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{La somme des coefficients de ce filtre étant non nulle, on sait que la couleur de l'image va intervenir après application du filtre.}$$

- un filtre (2) approchant la dérivée (∂_x) selon la direction horizontale x :

$$\frac{1}{6} \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad \text{Cette fois-ci la somme des coefficients est nulle, mais les coefficients de poids positifs et négatifs sont séparés en deux groupes, ce qui fait intervenir une dérivée première dans une direction.}$$

- autre exemple de dérivation, le filtre (3) :

$$\frac{1}{4\sqrt{2}} \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix} \quad \text{qui approche le dérivée dans la direction d'angle } \frac{\pi}{4} : \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right).$$

• un exemple (4) de filtres sont les filtres de Robinson [61] qui interviennent dans la détection de contours, il y a quatre filtres directionnels dont l'un d'eux est :

$$\frac{1}{5} \begin{pmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{pmatrix} \quad \text{On peut y reconnaître la combinaison linéaire entre la dérivée dans la direction d'angle } \frac{\pi}{4} \text{ et la dérivée seconde dans la direction d'angle } -\frac{\pi}{4}.$$

- le filtre (5) suivant:

$$\frac{1}{2} \begin{pmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \\ -1 & 0 & 1 \end{pmatrix} \quad \text{approche la dérivée troisième } \partial_x \partial_y^2.$$

La figure 3.2 montre l'action de chacun de ces filtres sur l'image 3.2.a.

Nous pouvons remarquer l'utilité que pourraient avoir certains filtres, par exemple pour la détection de contour avec les filtres (2), (3) et (4).

3.2.2 Définition

Un filtre s'appuie d'abord sur le voisinage d'un pixel, et on définit $n \in \mathbb{N}^*$, le rayon du filtre. En dimension 1, les voisinages centrés de rayons n sont $[-n, n]$, en 2 dimensions, les voisinages carrés centrés sont $[-n, n]^2$. Notons V l'ensemble des coordonnées du voisinage. On introduit le couple $k = (k_1, k_2)$ qui caractérise les coordonnées d'un point du voisinage autour d'un point X positionné en $I = (i, j)$. Un filtre h est alors une application qui donne la valeur du coefficient h_k du filtre en k .

Par exemple pour le filtre qui discrétise le laplacien,

$$h = \begin{pmatrix} h_{-1,1} & h_{0,1} & h_{1,1} \\ h_{-1,0} & h_{0,0} & h_{1,0} \\ h_{-1,-1} & h_{0,-1} & h_{1,-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad (3.3)$$

Un filtre linéaire est une application qui transforme une image en une autre, il se calcule en chaque pixel de l'image. La valeur de la nouvelle image après application du filtre au pixel positionné en $I = (i, j)$ est alors :

$$\sum_k h_k u_{I+k} \quad (3.4)$$

À un symbole $+$ ou $-$ près, on peut reconnaître la discrétisation du produit de convolution dans le cas continu :

$$(u * h)(X) = \int_Y h(Y) u(X - Y) dY \quad (3.5)$$

On peut alléger les notations en posant, pour un X fixé, $u_k = u(X + k)$ tel que :

$$\sum_k h_k u_{I+k} = \sum_k h_k u_k \quad (3.6)$$

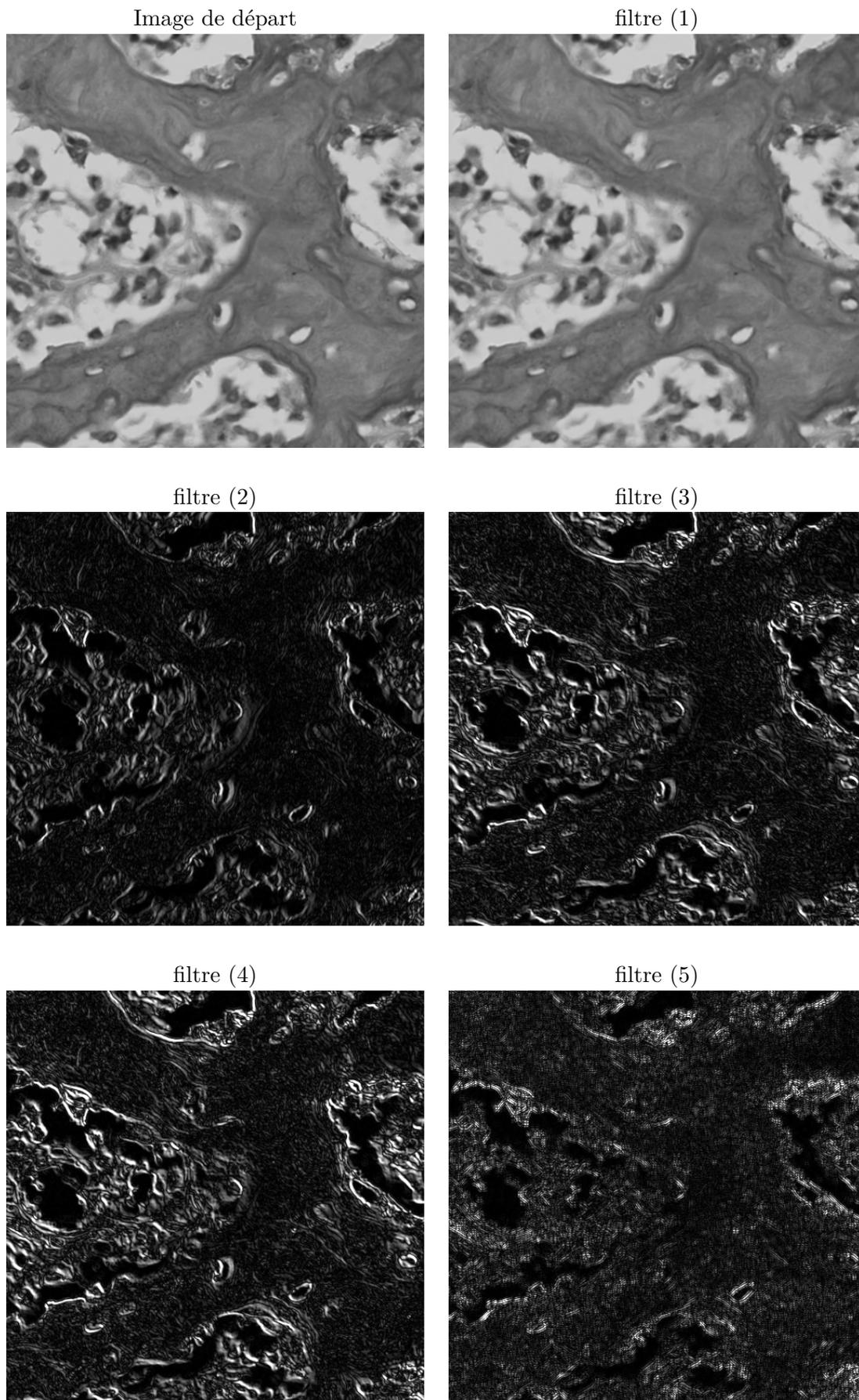


Figure 3.2: Exemple d'application des filtres linéaires 1 à 5 basés sur les dérivées partielles.

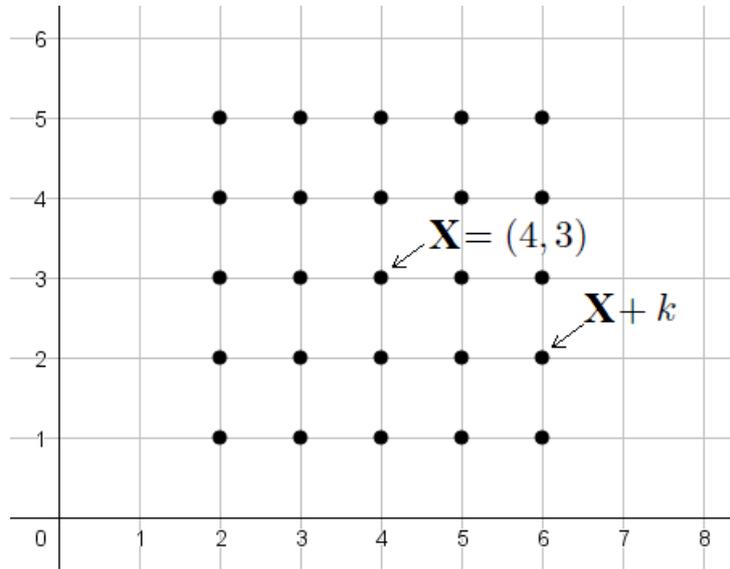


Figure 3.3: Pour cet exemple 2D, le voisinage est $V = [|-2, 2|]^2$, et la valeur de $k \in V$ montré dans le schema correspond à $k = (2, -1)$. Dans le cas de la dimension 2, $X = (x_1, x_2)$ et $k = (k_1, k_2)$ sont donc des couples.

A partir de là, 3 notations sont introduites:

- La notation multi-indices :

Dans le cas de la dimension 2, un multi-indice β est un élément de \mathbb{N}^2 , on pourrait l'expliquer en $\beta = (\beta_1, \beta_2)$.

Ainsi dans la suite, nous considérerons les notations suivantes :

$$X^\beta = x_1^{\beta_1} x_2^{\beta_2} \quad \partial_X^\beta u = \partial_{x_1}^{\beta_1} \partial_{x_2}^{\beta_2} u \quad \beta! = \beta_1! \beta_2! \quad |\beta| = \beta_1 + \beta_2 \quad (3.7)$$

- La notation des matrices indexées :

Nous pouvons expliciter les matrices avec les notations suivantes:

$$A = (a_{i,j})_{i,j} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & & & \vdots \\ \vdots & & \ddots & \\ a_{n,1} & \cdots & & a_{n,n} \end{pmatrix} \quad (3.8)$$

- La notation des segments d'entiers :

Le segment d'entiers $[1, n]$ désigne $[1, n] \cap \mathbb{Z}$.

3.2.3 Compréhension des filtres linéaires: exemple des filtres issus de la SVM

Les filtres linéaires issus de la littérature, très utiles pour le traitement des images, sont peu interprétables directement, ce qui peut devenir une limitation dans le traitement d'images en particulier lorsqu'il faut s'appuyer sur celui-ci dans la prise de décision médicale par exemple. Dans la suite, nous allons nous intéresser à la manière d'interpréter un filtre linéaire sur une image en tant que combinaison linéaire d'approximations de dérivées partielles, et en particulier les filtres issus de la SVM.

La SVM est un algorithme qui sépare deux groupes de points par une application linéaire [57]. Il ne prend pas en compte toutes les observations, mais se base sur les quelques points pertinents qui permettent de séparer les deux groupes. Dans le contexte des images étudiées dans cette thèse, les observations correspondent à l'ensemble des valeurs des couleurs sur un voisinage

de chaque pixel de l'image. Prendre en compte ce voisinage permet d'accéder localement à des informations de structure et de texture [62].

Soit $X = (i, j)$ la position d'un pixel et r_k , g_k et b_k les composantes rouges vertes et bleues de l'image en $X + k$, où k est dans le voisinage $[[-n, n]]$. Alors les observations sont les vecteurs colonnes $3(2n + 1)^2 \times 1$ en $c = ((r_k)_k, (g_k)_k, (b_k)_k)$.

On peut séparer le filtre linéaire en somme de trois filtres sur la couleur rouge, la couleur verte et la couleur bleue.

$$h(c) = h_r((r_k)_k) + h_g((g_k)_k) + h_b((b_k)_k) \quad (3.9)$$

La SVM fournit alors une application linéaire φ qui sépare deux ensembles de points échantillonnés sur des parties de l'image correspondant à de l'os ou à des cellules. Cette application linéaire permet en particulier de segmenter les cellules après seuillage (figure 3.4).

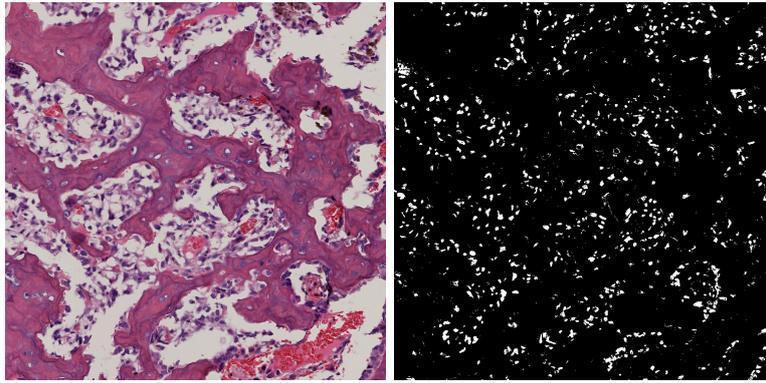


Figure 3.4: Application du filtre linéaire, partie positive en blanc et négatif en noir, classifiant les cellules en blanc.

En utilisant cette approche, nous obtenons le filtre suivant, à appliquer sur les trois composantes des couleurs RGB puis à sommer (figure 3.5).

-0.0322	-0.0018	-0.0095	-0.0034	-0.0450
0.0031	0.0192	0.0065	0.0239	0.0055
-0.0088	-0.0069	-0.0204	0.0028	-0.0033
-0.0000	0.0203	0.0106	0.0268	0.0003
-0.0299	0.0003	-0.0073	-0.0127	-0.0487
-0.0081	-0.0002	-0.0023	-0.0075	0.0188
-0.0026	0.0053	0.0037	-0.0139	0.0048
-0.0010	0.0093	0.0034	-0.0181	0.0023
-0.0103	-0.0030	-0.0102	-0.0290	-0.0009
-0.0044	0.0007	-0.0041	-0.0051	0.0602
0.0364	0.0066	0.0077	0.0094	0.0293
-0.0052	-0.0208	-0.0090	-0.0140	-0.0113
0.0110	-0.0058	0.0131	0.0108	0.0127
0.0056	-0.0149	0.0013	-0.0017	-0.0097
0.0321	-0.0021	0.0126	0.0050	0.0110

Figure 3.5: Exemple de filtre linéaire obtenu par SVM et appliqué sur les composantes rouge, verte et bleue de données issues d'une coupe histologique.

Nous comprenons que ces filtres sont difficilement interprétables sans méthode particulière [63]. Dans la suite, une méthode permettant de décomposer un tel filtre en approximations de

dérivées partielles est établie. Dans un premier temps, la décomposition se fera en considérant l'image comme non bruitée puis bruitée.

Si quelques références de la littérature s'intéressent plus ou moins à cette question [64, 65, 66], en particulier en lien avec la décomposition des images en structure et texture via des filtres de Gabor [67] ou autres stratégies [68], à notre connaissance, aucune proposition d'approche générale comme celle explicitée ici n'a été faite, en particulier en ce qui concerne le lien entre les filtres issus de la SVM et des filtres pouvant être calculés analytiquement. Majoritairement les références de la littérature s'intéressent à une décomposition de l'image en ondelettes en séparant structure et texture.

3.3 Décomposition d'un filtre linéaire sur un signal "lisse".

L'hypothèse principale dans cette section, est de considérer des images u non-bruitées.

3.3.1 Approximation de la dérivée partielle ∂^γ .

Dans un premier temps, nous allons estimer la dérivée partielle de multi-indice γ en un pixel de l'image à partir des valeurs sur son voisinage. Chaque étape sera illustrée, dans un encadré bleu, sur un exemple concret.

Soit $n \in \mathbb{N}^*$ la largeur de la fenêtre avec $V = \llbracket -n, n \rrbracket^2$ l'ensemble des voisinages, et soit $\beta \in W = \llbracket 0, 2n \rrbracket^2$ les multi-indices qui interviennent dans le développement limité de u .

Les calculs présentés dans cette partie considèrent $n = 1$, c'est à dire $V = \llbracket -1, 1 \rrbracket^2$ et $W = \llbracket 0, 2 \rrbracket^2$.

Le voisinage est choisi volontairement petit pour pouvoir effectuer aisément les calculs analytiques cependant des voisinages de tailles plus importantes pourront être développés dans la suite.

On approxime u sur le voisinage $X + k$ de X où $k \in V$ avec son développement limité:

$$u(X + k) = u_k = \sum_{\beta \in W} \frac{1}{\beta!} \partial_X^\beta u(X) k^\beta + \text{résidu} \quad (3.10)$$

On cherche la combinaison linéaire des u_k qui approche la dérivée ∂_X^γ , avec α_k ses coefficients et $\gamma \in W$. On peut les noter avec un exposant γ pour se souvenir qu'ils approximent ∂^γ . Ainsi le coefficient en $k \in V$ se note α_k^γ .

$$\sum_{k \in V} \alpha_k^\gamma u_k = \sum_{k \in V} \alpha_k^\gamma \sum_{\beta \in W} \frac{k^\beta}{\beta!} \partial_X^\beta u(X) = \sum_{\beta \in W} \left(\sum_{k \in V} \frac{k^\beta}{\beta!} \alpha_k^\gamma \right) \partial_X^\beta u(X) \quad (3.11)$$

Nous obtenons une combinaison linéaire sur les dérivées $\partial_X^\beta u(X)$. De manière à obtenir $\sum_k \alpha_k^\gamma u_k \approx \partial_X^\gamma u(X)$, on cherche les α_k^γ tels que :

$$\forall \beta \in W \quad \sum_{k \in V} \frac{k^\beta}{\beta!} \alpha_k^\gamma = \begin{cases} 0 & \text{si } \beta \neq \gamma \\ 1 & \text{si } \beta = \gamma \end{cases} \quad (3.12)$$

C'est un système linéaire d'inconnues les coefficients α_k^γ , indexé par $k \in \llbracket -n, n \rrbracket^2$, et où chaque ligne correspondent à un $\beta \in \llbracket 0, 2n \rrbracket^2$.

On peut ordonner arbitrairement les k et les β de manière à exprimer ce système matriciellement.

D'abord, on choisit d'ordonner les k et β de la manière suivante :

k_1	-1	0	1	-1	0	1	-1	0	1
k_2	-1	-1	-1	0	0	0	1	1	1

Figure 3.6: Tableau ordonnant les valeurs $k = (k_1, k_2)$ de V

β_1	0	1	2	0	1	2	0	1	2
β_2	0	0	0	1	1	1	2	2	2

Figure 3.7: Tableau ordonnant les valeurs $\beta = (\beta_1, \beta_2)$ de W

Ces tableaux donnent l'ordre pour les valeurs de k et β , nous nous reporterons à ces tableaux lorsque nous devons lire des vecteurs indexés par k ou β .

Soient la matrice A et le vecteur α^γ tels que :

$$A = \left(\frac{k^\beta}{\beta!} \right)_{\beta, k} \quad \alpha^\gamma = (\alpha_k^\gamma)_k \quad (3.13)$$

On peut remarquer que la matrice A est similaire à une matrice de Vandermonde à un coefficient $\frac{1}{\beta!}$ près, dans une version à plusieurs variables.

Nous pouvons alors calculer la matrice $A = \left(\frac{k^\beta}{\beta!} \right)_{\beta, k}$ sachant que $\frac{k^\beta}{\beta!} = \frac{k_1^{\beta_1}}{\beta_1!} \cdot \frac{k_2^{\beta_2}}{\beta_2!}$:

β	k	-1	0	1	-1	0	1	-1	0	1
		-1	-1	-1	0	0	0	1	1	1
0	0	1	1	1	1	1	1	1	1	1
1	0	-1	0	1	-1	0	1	-1	0	1
2	0	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
0	1	-1	-1	-1	0	0	0	1	1	1
1	1	1	0	-1	0	0	0	-1	0	1
2	1	$-\frac{1}{2}$	0	$-\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
0	2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
1	2	$-\frac{1}{2}$	0	$\frac{1}{2}$	0	0	0	$-\frac{1}{2}$	0	$\frac{1}{2}$
2	2	$\frac{1}{4}$	0	$\frac{1}{4}$	0	0	0	$\frac{1}{4}$	0	$\frac{1}{4}$

La matrice A est donc :

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} \\ -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 \\ -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} \end{pmatrix} \quad (3.14)$$

Soit le vecteur colonne e_γ qui vaut 1 en γ et 0 partout ailleurs :

$$e_\gamma = (0, \dots, 0, 1, 0, \dots, 0)^T \quad \text{ou} \quad e_\gamma = (\delta_\gamma(\beta))_\beta \quad \text{avec symbole de Kronecker} \quad (3.15)$$

L'ensemble $\{e_\gamma, \gamma \in W\}$ forme la base canonique de l'espace vectoriel.

Nous avons alors, d'après l'équation (3.12), l'expression matricielle du système :

$$A\alpha^\gamma = e_\gamma \quad (3.16)$$

Les coefficients recherchés sont donc :

$$\alpha^\gamma = A^{-1}e_\gamma \quad (3.17)$$

L'inverse de la matrice A est :

$$A^{-1} = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{4} & -\frac{1}{2} & 0 & -\frac{1}{2} & 1 \\ 0 & 0 & 0 & -\frac{1}{2} & 0 & 1 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & -\frac{1}{4} & -\frac{1}{2} & 0 & \frac{1}{2} & 1 \\ 0 & -\frac{1}{2} & 1 & 0 & 0 & 0 & 0 & 1 & -2 \\ 1 & 0 & -2 & 0 & 0 & 0 & -2 & 0 & 4 \\ 0 & \frac{1}{2} & 1 & 0 & 0 & 0 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 & -\frac{1}{4} & \frac{1}{2} & 0 & -\frac{1}{2} & 1 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & -1 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{2} & 1 \end{pmatrix} \quad (3.18)$$

On en déduit alors que A^{-1} est une matrice qui contient tous les coefficients :

$$(\alpha_k^\gamma)_{k,\gamma} = A^{-1} \quad (3.19)$$

Concrètement, α^γ est le vecteur dans lequel on trouve les coefficients à mettre dans le filtre recherché approximant la dérivée partielle $\partial_x^\gamma u(x)$.

On sait que $A^{-1} = (\alpha_k^\gamma)_{k,\gamma}$.

On lit alors les valeurs du filtre α^γ sur la " γ -ième" colonne de A^{-1} .

Par exemple, si on veut connaître le filtre correspondant à la dérivée $\partial_x^2 \partial_y$, on cherche la colonne associée à $\gamma = (2, 1)$. Elle correspond à la 6^{ème} d'après notre tableau qui ordonne les β .

Nous lisons le vecteur colonne :

$$(\alpha_k^{(2,1)})_k = \begin{pmatrix} -\frac{1}{2} \\ 1 \\ -\frac{1}{2} \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{2} \\ -1 \\ \frac{1}{2} \end{pmatrix} \quad (3.20)$$

Chaque coefficient de ce vecteur colonne est le coefficient $\alpha_k^{(2,1)}$ associé au k correspondant. On lit le tableau qui ordonne k est on peut reconstituer le filtre :

$$h = \frac{1}{2} \begin{pmatrix} 1 & -2 & 1 \\ 0 & 0 & 0 \\ -1 & 2 & -1 \end{pmatrix} \quad (3.21)$$

Ainsi:

$$\sum_k \alpha_k^\gamma u_k \approx \partial^\gamma u(x) \quad (3.22)$$

où α^γ est alors le filtre qui approxime la dérivée partielle ∂^γ .

Dans le contexte d'une image non bruitée, le calcul de tous les filtres α^γ se font simultanément en inversant la matrice A .

Voici maintenant une illustration d'un de ces filtres pour une valeur de n fixée à 3 pour un voisinage $[[-3, 3]]^2$ dans le cas 2D. La matrice A^{-1} calculée est la matrice qui contient tous les filtres en vecteur colonne. Avec ce choix d'ordonnement des β , le numéro de la colonne correspondant à $\gamma = (\gamma_1, \gamma_2)$ est : $1 + \gamma_1 + (2 * n + 1) * \gamma_2$.

Sur la figure 3.8, on visualise le filtre linéaire associé à la dérivée partielle $\partial_x^3 \partial_y^3$, sous différentes représentations, c'est-à-dire, nuage de points, surface et diagramme à barres. . Chaque point a pour coordonnées $(k_1, k_2, \alpha_k^{(3,3)})$.

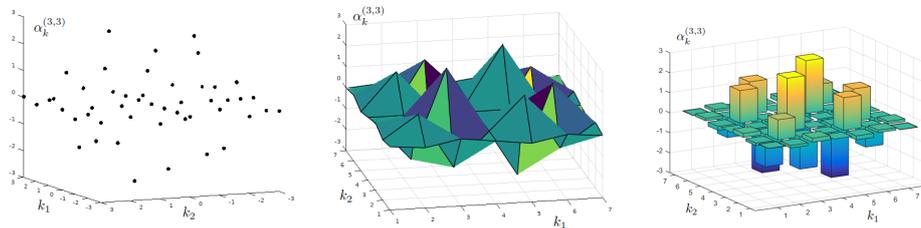
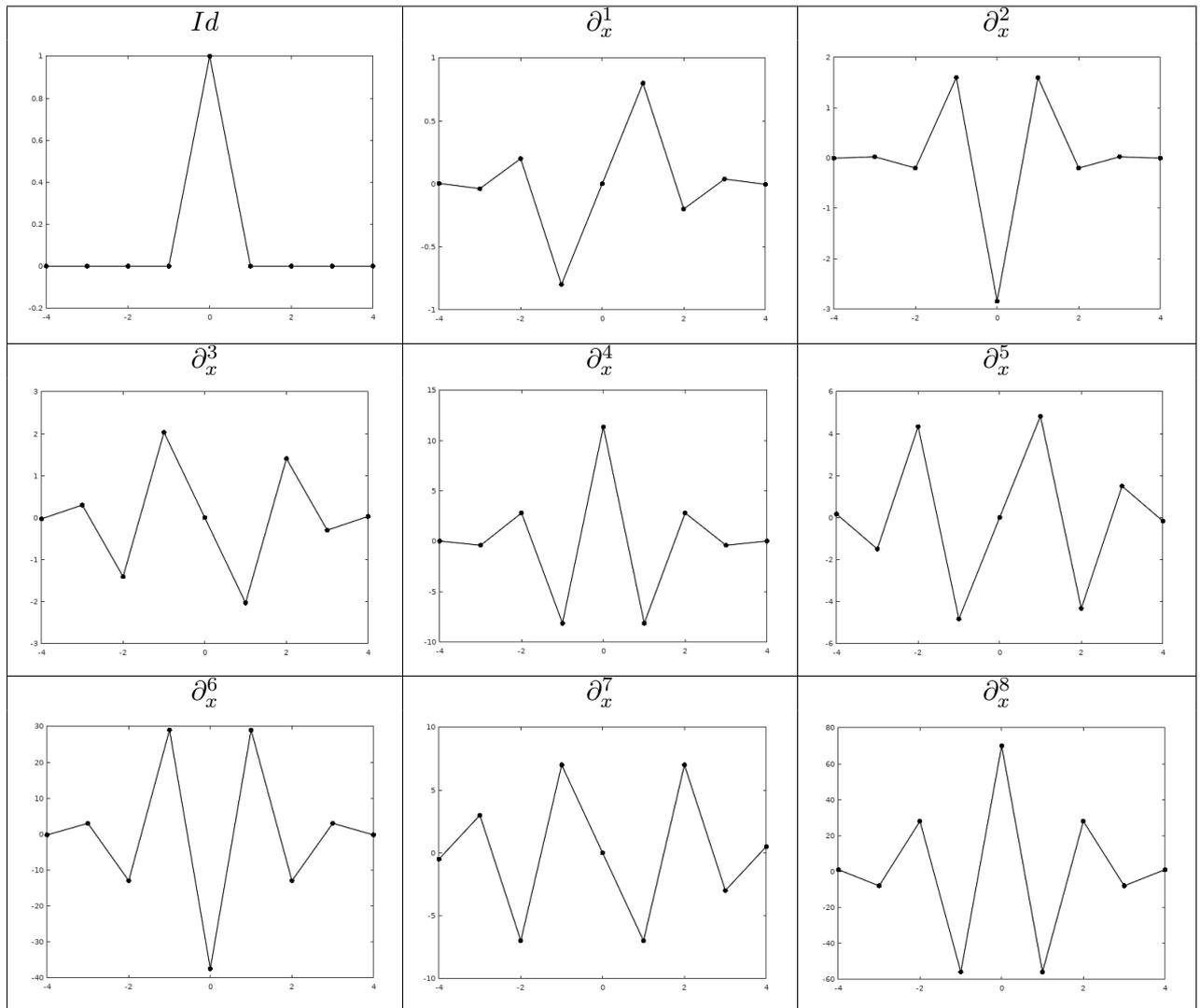


Figure 3.8: Représentations du filtre linéaire associé à la dérivée partielle $\partial_x^3 \partial_y^3$ en nuage de points, surface et diagramme à barre.

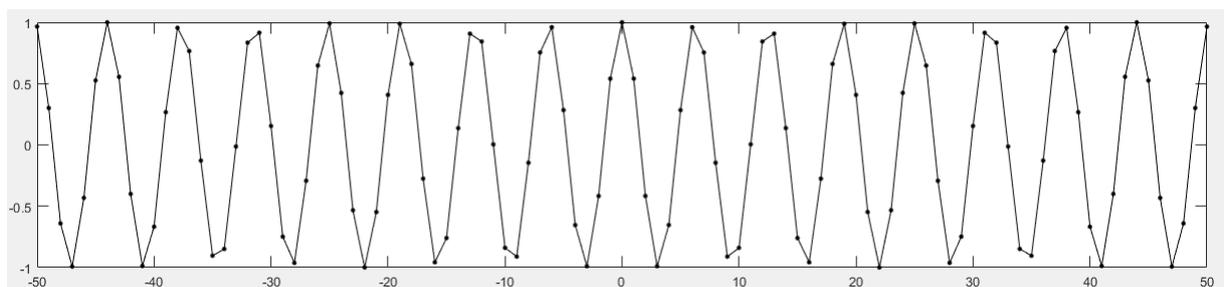
3.3.2 Exemple de l'approximation des filtres et comparaison avec les différences finies dans le cas 1D

Dans un but illustratif et de validation de l'approche, nous allons nous intéresser à un signal 1D. La figure 3.9 montre les profils des filtres obtenus pour un voisinage en $[[-4, 4]]$:

Figure 3.9: Filtres 1D avec $V =]-4, 4[$.

Remarquons que le filtre α^0 , qui correspond à l'identité, ne s'appuie que sur la valeur u_0 . Il est logique d'obtenir ce résultat pour un signal non bruité. Cependant, un tel filtre sera très sensible au bruit et nous discuterons ce point dans la suite de ce chapitre.

Une stratégie usuelle numérique pour approximer des dérivées partielles est de s'appuyer sur la méthode des différences finies. En développant cette approche d'approximation de dérivées partielles par des filtres, on peut tester leur qualité en comparaison avec l'approche différences finies. Dans un premier temps, on s'intéresse à la fonction cosinus discrétisée sur les entiers appartenant à $[-50, 50]$ (figure 3.10). Cette discrétisation induit une perte d'information et on cherche à savoir si les filtres sont capables de retrouver avec précision les dérivées analytiques de cette fonction à partir du signal discret.

Figure 3.10: Fonction cosinus sur $[-50, 50]$.

Nous allons maintenant appliquer le filtre α^1 qui approxime la dérivée d'ordre 1, puis superposer les courbes de la dérivée analytique ($-\sin(x)$) et celle obtenue par différence finie sur un maillage grossier. La figure 3.11 montre en noir la dérivée analytique, en vert l'approximation avec le filtre α^1 et en bleu celle des différences finies.

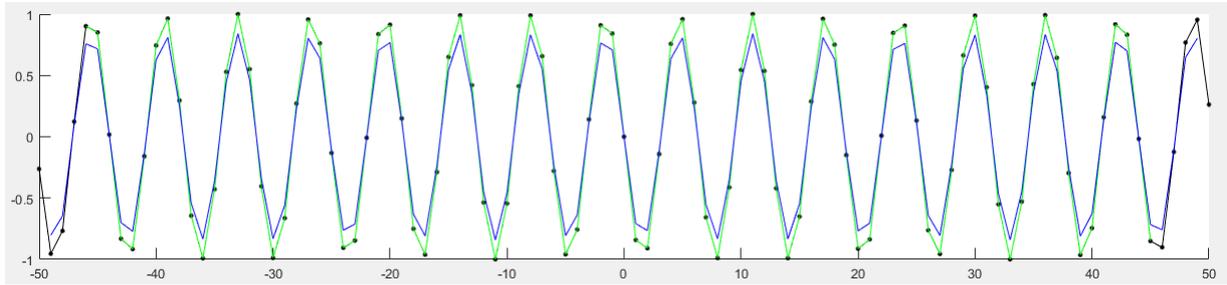


Figure 3.11: En noir : dérivée analytique, en bleu : différences finies, en vert : application du filtre α^1

Les courbes verte et noire sont superposées, ce qui montre que le filtre α_1 approche beaucoup mieux la dérivée sur la fonction cosinus que l'approximation de celle-ci par différences finies. Quantitativement, on peut calculer les carrés des normes L^2 des différences suivantes:

$$\|v_{\text{filtre}} - v_{\text{derivee_analytique}}\|_{L^2}^2 \approx 6,75 \times 10^{-5} \quad \|v_{\text{diff_finies}} - v_{\text{derivee_analytique}}\|_{L^2}^2 \approx 1,16 \quad (3.23)$$

Le tableau 3.12 indique les carrés des normes L^2 obtenues pour les dérivées 1, 2, 3 et 4 :

$\gamma =$	1	2	3	4
$\ v_{\text{filtre}} - v_{\text{derivee_analytique}}\ _{L^2}^2$	$6,75 \times 10^{-5}$	$2,93 \times 10^{-6}$	5×10^{-3}	$8,64 \times 10^{-4}$
$\ v_{\text{diff_finies}} - v_{\text{derivee_analytique}}\ _{L^2}^2$	1,16	$2,97 \times 10^{-1}$	2,37	1,10

Figure 3.12: Comparaison entre les différences finies et les filtres α^k des carrés des normes L^2 de différences avec la dérivée analytique.

On peut continuer ainsi, on observe une différence de précision qui diminue, entre les filtres calculés par la méthode proposée et ceux obtenus par différences finies. Pour les deux dernières dérivées (7^{ième} et 8^{ième}), les filtres α^7 et α^8 génèrent les mêmes erreurs que la méthode aux différences finies, figure 3.13:

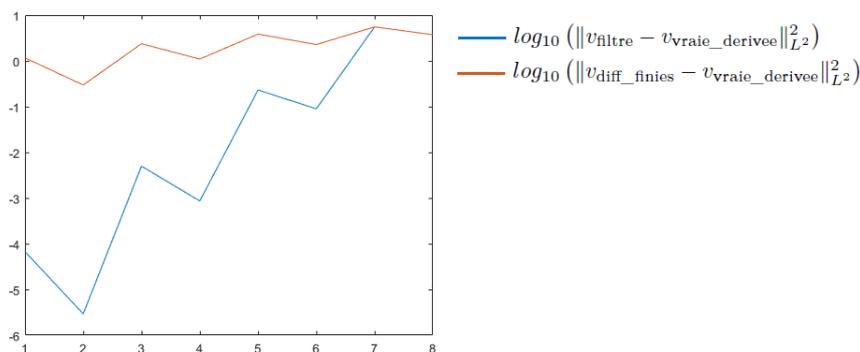


Figure 3.13: comparaison des distances $\| \cdot \|_{L^2}^2$ avec échelle logarithmique

Les filtres α^γ donnent de meilleurs résultats, aux petits ordres, que ceux issus des différences finies dans ce cas particulier de l'étude de la fonction cosinus.

3.3.3 Décomposition d'un filtre

Nous avons vu que nous pouvions construire des filtres approchant des dérivées partielles, mais le fait d'avoir déterminé la matrice A nous permet aussi de décomposer un filtre donné.

Étant donné un filtre quelconque $h = (h_k)_k$, on peut le décomposer en combinaison linéaire de filtres α^γ .

Soit h tel que :

$$h = \sum_{\gamma} \lambda_{\gamma} \alpha^{\gamma} \quad (3.24)$$

On cherche les coefficients λ_{γ} . Alors, grâce à l'équation 3.16 :

$$Ah = A \cdot \sum_{\gamma} \lambda_{\gamma} \alpha_k^{\gamma} = \sum_{\gamma} \lambda_{\gamma} e_{\gamma} = (\lambda_{\gamma})_{\gamma} \quad (3.25)$$

Donc, en se donnant un filtre h et connaissant A , on peut déterminer quels sont les coefficients λ_{γ} . Le vecteur λ donne alors les contributions (algébriquement) de chaque dérivées dans le filtre initial h .

La matrice A est alors une matrice de changement de base entre la base canonique des filtres et la base $(\alpha^{\gamma})_{\gamma}$ des approximations des dérivées partielles.

Pour illustrer ce point, soit h un des filtres directionnels de Kirch [61] qui intervient dans la détection de contour:

$$h = \frac{1}{15} \begin{pmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{pmatrix} \quad (3.26)$$

Nous avons déjà déterminé la matrice A pour un voisinage 3×3 en $[-1, 1]^2$ dans la section 3.3.1. En linéarisant h et en appliquant A , on peut alors déterminer quelles dérivées partielles interviennent dans ce filtre, n lit les valeurs λ_{γ} dans l'ordre choisi pour les valeurs de β .

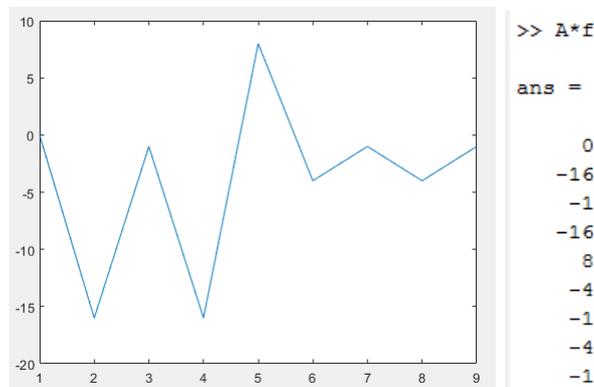


Figure 3.14: Coefficients λ_{γ} de la combinaison linéaire $h = \sum \lambda_{\gamma} \alpha^{\gamma}$

Nous pouvons alors donner le pourcentage des dérivées partielles qui interviennent dans ce filtre et extraire les contributions de chaque dérivées partielles (tableau 3.1).

On voit ici que les dérivées impliquant une dérivée d'ordre impair ont des contributions plus importantes corroborant l'aspect directionnel du filtre de Kirch ci-dessus.

Dérivées partielle	Id	∂_x	∂_y	∂_x^2	$\partial_x \partial_y$	∂_y^2	$\partial_x^2 \partial_y$	$\partial_x \partial_y^2$	$\partial_x^2 \partial_y^2$
Valeur	0	-16	-16	-1	8	-1	-4	-4	-1
Pourcentage	0%	31,37%	31,37%	1,96%	15,69%	1,96%	7,84%	7,84%	1,96%

Table 3.1: Tableau des valeurs et pourcentage des coefficients λ_{γ} dans le filtre directionnel de Kirch.

3.4 Décomposition d'un filtre linéaire sur un signal bruité

Les filtres permettant d'approcher des dérivées partielles sur des signaux non bruités peuvent se révéler être très sensibles au bruit. Une manière d'étudier une image bruitée par des filtres établis sur la base de dérivées partielles est d'appliquer d'abord un flou gaussien, avec une largeur adaptée, puis d'appliquer les filtres linéaires choisis. La question qui découle est alors de savoir comment déterminer des filtres linéaires qui prennent directement en compte le bruit d'une image.

3.4.1 Approximation de la dérivée partielle ∂^γ .

On utilise les mêmes notations que dans la section précédente où le bruit n'est pas pris en compte avec $n \in \mathbb{N}^*$, $\beta \in [0, 2n]^2$, $k \in [-n, n]^2$.

$$u(X + k) = u_k = \sum_{\beta} \frac{1}{\beta!} \partial_X^\beta u(X) k^\beta \quad (3.27)$$

$$v_k = u(X + k) + \varepsilon_k \quad \varepsilon_k \sim \mathcal{N}(0, \sigma^2) \text{ iid} \quad (3.28)$$

Ici, v_k est l'image bruitée où ε_k joue le rôle du bruit et correspond à une distribution normale de moyenne nulle et d'écart type σ . En pratique, nous n'avons accès qu'à v_k , l'image bruitée. Dans le cadre de cette thèse, celles-ci sont obtenues par des techniques d'anatomo-pathologie sur des patients atteints d'ostéosarcome.

Comme précédemment, nous avons la combinaison linéaire :

$$\sum_k \alpha_k u(X + k) = \sum_k \alpha_k \sum_{\beta} \frac{1}{\beta!} \partial_X^\beta u(X) k^\beta = \sum_{\beta} \left(\sum_k \frac{k^\beta}{\beta!} \alpha_k \right) \partial_X^\beta u(X) \quad (3.29)$$

En prenant en compte le bruit, nous obtenons :

$$\sum_k \alpha_k v_k = \sum_{\beta} \left(\sum_k \frac{k^\beta}{\beta!} \alpha_k \right) \partial_X^\beta u(X) + \sum_k \alpha_k \varepsilon_k \quad (3.30)$$

On cherche un estimateur de $\partial_x^\gamma u(x)$. Cela revient à déterminer les coefficients α_k^γ qui vont minimiser l'erreur :

$$\mathbb{E} \left(\left| \sum_k \alpha_k v_k - \partial_X^\gamma u(X) \right|^2 \right) = \mathbb{E} \left(\underbrace{\left| \sum_k \alpha_k v_k - \mathbb{E} \left(\sum_k \alpha_k v_k \right) \right|^2}_{\text{(I)}} + \underbrace{\left| \mathbb{E} \left(\sum_k \alpha_k v_k \right) - \partial_X^\gamma u(X) \right|^2}_{\text{(III)}} \right) \quad (3.31)$$

On est donc mené à faire le calcul habituel du dilemme biais-variance [57].

$$\text{(I)} = \mathbb{E} \left(\sum_k \alpha_k u(X + k) + \sum_k \alpha_k \varepsilon_k \right) = \sum_k \alpha_k u(X + k) \quad (3.32)$$

$$\text{(II)} = \sum_k \alpha_k v_k - \sum_k \alpha_k u(X + k) = \sum_k \alpha_k \varepsilon_k \quad (3.33)$$

$$\text{(III)} = \sum_k \alpha_k u_k - \partial_X^\gamma u(X) = \sum_{\beta} \left(\sum_k \frac{k^\beta}{\beta!} \alpha_k \right) \partial_X^\beta u(X) - \partial_X^\gamma u(X) \quad (3.34)$$

(II) est d'espérance nulle, car les ε_k sont centrés et (III) ne fait plus intervenir d'aléatoire. Le calcul biais-variance est alors :

$$\mathbb{E}\left(\left|\sum_k \alpha_k v_k - \partial_x^\gamma u(x)\right|^2\right) = \mathbb{E}\left((3.33)^2\right) + 2\mathbb{E}\left((3.33) \times (3.34)\right) + \mathbb{E}\left((3.34)^2\right) = \mathbb{E}\left((3.33)^2\right) + (3.34)^2 \quad (3.35)$$

$$= \underbrace{\mathbb{E}\left(\left|\sum_k \alpha_k \varepsilon_k\right|^2\right)}_{(i)} + \underbrace{\left|\sum_\beta \left(\sum_k \frac{k^\beta}{\beta!} \alpha_k\right) \partial_X^\beta u(X) - \partial_X^\gamma u(X)\right|^2}_{(ii)} \quad (3.36)$$

$$(i) = \text{Var}\left(\sum_k \alpha_k \varepsilon_k\right) = \sigma^2 \sum_k \alpha_k^2 \quad (3.37)$$

$$(ii) = \left(\sum_\beta \left(\sum_k \frac{k^\beta}{\beta!} \alpha_k\right) \partial_X^\beta u(X)\right)^2 - 2 \sum_\beta \left(\sum_k \frac{k^\beta}{\beta!} \alpha_k\right) \partial_X^\beta u(X) \partial_X^\gamma u(X) + (\partial_X^\gamma u(X))^2 \quad (3.38)$$

Soit ϕ l'erreur :

$$\phi(\alpha^\gamma) = (3.37) + (3.38) \quad (3.39)$$

ϕ est une forme quadratique de α^γ .

On veut minimiser ϕ , on cherche alors α^γ telle que:

$$\frac{1}{2} \nabla_{\alpha^\gamma} \phi(\alpha^\gamma) = 0 \quad (3.40)$$

$$\Leftrightarrow \sigma^2 \alpha^\gamma + \sum_\beta \left(\sum_k \frac{k^\beta}{\beta!} \alpha_k\right) \partial_X^\beta u(X) \sum_\eta \left(\sum_m \frac{k^\eta}{\eta!} e_m\right) \partial_X^\eta u(X) - \sum_\eta \left(\sum_m \frac{m^\eta}{\eta!} e_m\right) \partial_X^\eta u(X) \partial_X^\gamma u(X) = 0 \quad (3.41)$$

$$\Leftrightarrow \forall m, \quad \sigma^2 \alpha_m^\gamma + \sum_k \left(\sum_{\beta, \eta} \frac{k^\beta m^\eta}{\beta! \eta!} \partial_X^\beta u(X) \partial_X^\eta u(X)\right) \alpha_k^\gamma - \sum_\eta \frac{m^\eta}{\eta!} \partial_X^\eta u(X) \partial_X^\gamma u(X) = 0 \quad (3.42)$$

Soit $\tilde{A}(u)$ la matrice :

$$\tilde{A}(u) = \sigma^2 I_{(2n+1)^2} + \left(\sum_{\beta, \eta} \frac{k^\beta m^\eta}{\beta! \eta!} \partial_X^\beta u(X) \partial_X^\eta u(X)\right)_{m, k} \quad (3.43)$$

Et $\tilde{C}_\gamma(u)$ le vecteur colonne :

$$\tilde{C}_\gamma(u) = \left(\sum_\eta \frac{m^\eta}{\eta!} \partial_X^\eta u(X) \partial_X^\gamma u(X)\right)_m \quad (3.44)$$

Dans un premier temps, on peut supposer que u et ses dérivées partielles sont des processus gaussiens stationnaires indépendants, de fonction moyenne nulle et de fonction de covariance K .

On cherche l'espérance de $\tilde{A}(u)$ et $\tilde{C}_\gamma(u)$ où la variable aléatoire est la fonction u .

$$\mathbb{E}(\tilde{A}(u)) = \sigma^2 I_{(2n+1)^2} + \left(\sum_{\beta, \eta} \frac{k^\beta m^\eta}{\beta! \eta!} \text{Cov}\left(\partial_X^\beta u(X), \partial_X^\eta u(X)\right)\right)_{m, k} \quad (3.45)$$

On a :

$$\text{Cov}\left(\partial_X^\beta u(X), \partial_X^\eta u(X)\right) = \begin{cases} K(0) & \text{si } \beta = \eta \\ 0 & \text{si } \beta \neq \eta \end{cases} \quad (3.46)$$

Donc :

$$\tilde{A} = \mathbb{E}(\tilde{A}(u)) = \sigma^2 I_{(2n+1)^2} + K(0) \left(\sum_{\beta} \frac{(km)^\beta}{(\beta!)^2} \right)_{m,k} \quad (3.47)$$

De même :

$$\tilde{C}_\gamma = \mathbb{E}(\tilde{C}_\gamma(u)) = K(0) \left(\frac{m^\gamma}{\gamma!} \right)_m \quad (3.48)$$

On a :

$$\tilde{A}\alpha^\gamma = \tilde{C}_\gamma \quad (3.49)$$

On peut diviser par $K(0)$,

$$A = \frac{\tilde{A}}{K(0)} \quad C_\gamma = \frac{\tilde{C}_\gamma}{K(0)} \quad (3.50)$$

On a alors :

$$A\alpha^\gamma = C_\gamma \quad (3.51)$$

avec :

$$A = \frac{\sigma^2}{K(0)} I_{(2n+1)^2} + \left(\sum_{\beta} \frac{(km)^\beta}{(\beta!)^2} \right)_{m,k} \quad (3.52)$$

$$C_\gamma = \left(\frac{m^\gamma}{\gamma!} \right)_m \quad (3.53)$$

Le terme en $\frac{\sigma^2}{K(0)}$ montre que l'écart-type du bruit sur l'image n'intervient que relativement à l'écart-type sur les couleurs de l'image ($K(0) = \text{Var}(u(X))$).

On peut donc déterminer le filtre α^γ qui approche la dérivée partielle $\partial_x^\gamma u(x)$:

$$\alpha^\gamma = A^{-1}C_\gamma \quad (3.54)$$

Soit C la matrice composée des vecteurs colonnes C_γ :

$$C = (C_\gamma)_\gamma \quad (3.55)$$

Si on souhaite avoir les coefficients α_k^γ dans une matrice :

$$(\alpha_k^\gamma)_{k,\gamma} = A^{-1}C \quad (3.56)$$

Nous pouvons, suite à ce premier calcul, chercher une estimation de $\text{Cov}\left(\partial_X^\beta u(X), \partial_X^\eta u(X)\right)$ plus adaptative aux types d'images étudiées que l'approximation de l'équation 3.46.

Reprenons les filtres obtenus avec l'approximation 3.56, nommons les ξ^γ . Nous avons l'approximation de la dérivée partielle associé à γ :

$$\partial^\gamma u(X) \approx \sum_k \xi_k^\gamma u_k \quad (3.57)$$

$$\text{Cov} \left(\partial_X^\beta u(X), \partial_X^\eta u(X) \right) = \text{Cov} \left(\sum_k \xi_k^\beta u_k, \sum_p \xi_p^\eta u_p \right) \quad (3.58)$$

$$= \sum_{k,p} \xi_k^\beta \text{Cov}(u_k, u_p) \xi_p^\eta = (\xi^\beta)^T (\text{Cov}(u_k, u_p))_{k,p} \xi^\eta \quad (3.59)$$

En tant que processus gaussien, on retrouve la fonction de covariance :

$$\text{Cov}(u_k, u_p) = K(p - k) \quad (3.60)$$

Ainsi, on peut estimer le terme avec les covariance en :

$$\text{Cov} \left(\partial_x^\beta u(x), \partial_x^\eta u(x) \right) = (\xi^\beta)^T \Gamma \xi^\eta \quad (3.61)$$

où Γ est la matrice de covariance $(K(p - k))_{k,p}$.

Par cette approximation plus détaillée de $\text{Cov} \left(\partial_X^\beta u(X), \partial_X^\eta u(X) \right)$, les filtres obtenus serait mieux adaptés à la famille d'images étudiée (images lisses, irrégulières, ...)

3.4.2 Décomposition d'un filtre sur un signal bruité

Nous avons :

$$A\alpha^\gamma = Ce_\gamma \quad (3.62)$$

$$\implies C^{-1}A\alpha^\gamma = e_\gamma \quad (3.63)$$

De même que dans la partie précédente dans le cas non bruitée, si on se donne un filtre h avec :

$$h = \sum_\gamma \lambda_\gamma \alpha^\gamma \quad (3.64)$$

On a :

$$C^{-1}Ah = C^{-1}A \sum_\gamma \lambda_\gamma \alpha^\gamma = \sum_\gamma \lambda_\gamma e_\gamma = (\lambda_\gamma)_\gamma \quad (3.65)$$

Le vecteur λ donne les contributions (algébriquement) de chaque dérivées dans le filtre initial h .

Cette approche rappelle l'approche proposée par Wiener pour des filtres du même nom [69].

En effet, un filtre de Wiener est un filtre linéaire et un estimateur (au sens statistique), obtenu par minimisation de l'erreur quadratique. Le signal de départ peut être discret ou continu, il comporte un bruit et la quantité à estimer peut être, par exemple, le futur d'une série temporelle, un signal (image, son, électrique, mécanique...) débruité.

Les filtres de Wiener sont utilisés dans le débruitage d'images [70]. Ce débruitage se fait dans l'espace de Fourier. Soit u l'image à reconstruire, g une transformation linéaire, ϵ un bruit et v l'image observée telle que:

$$v = u * g + \epsilon \quad (3.66)$$

On cherche à reconstruire u avec un filtre h . Soit \hat{u} l'estimateur de u .

$$\hat{u} = v * h \quad (3.67)$$

Soient G la transformée de fourier de g , H la transformée de h , S la densité spectrale de puissance moyenne de u et N la densité spectrale de puissance moyenne de ϵ .

Alors la transformée de fourier de h est :

$$H = \frac{G^* S}{|G|^2 S + N} \quad (3.68)$$

Dans cette thèse, les filtres développés sont considérés comme des estimateurs linéaires de dérivées partielles pour un signal bruité, et sont obtenus par minimisation de l'erreur quadratique et en ce sens se rapprochent des théories proposées par Wiener.

3.5 Opérateurs différentiels invariants par rotation et symétrie

3.5.1 Introduction

Lorsqu'on s'intéresse aux profils des filtres permettant de séparer les cellules de l'os en les décomposant, on observe que certaines dérivées partielles telles que ∂_x et ∂_y ont des contributions quasi-nulles. D'autre part, ∂_x et ∂_y sont des dérivées très directionnelles (filtres directionnels [66]). Étant donné que l'orientation de l'image de la coupe histologique est arbitraire, la segmentation que l'on souhaite obtenir doit être invariante par rotation. C'est une des raisons possibles pour laquelle seuls les opérateurs différentiels invariants par rotation contribuent significativement aux filtres pertinents pour l'analyse des images de cette thèse.

Passer dans l'espace des approximations des dérivées partielles à partir de la base canonique des filtres correspond à un changement de base et il suffit d'appliquer la matrice A^{-1} ou $A^{-1}C$ des sections 3.3.1 et 3.4.1 pour l'obtenir. La décomposition des filtres issus de la SVM sur la base des filtres aux dérivées partielles permet de réduire la dimension de l'étude en se restreignant à des sous-espaces pertinents. De plus, comme nous sommes amenés à changer de base et donc à avoir les valeurs des approximations des dérivées partielles, nous pouvons en profiter pour y inclure des opérateurs non-linéaires. En effet certaines quantités comme, la norme du gradient, sont des informations intéressantes pour le traitement d'images.

Pour être robuste à la rotation d'une image, nous allons chercher des opérateurs différentiels invariants par rotation. Dans la suite de nouvelles notations seront utilisées relativement aux sections précédentes.

Soient $f : \mathbb{R}^2 \mapsto \mathbb{R} C^\infty$, et φ , une fonction sur les dérivées partielles $\partial_x f, \partial_y f, \partial_x^2 f, \dots$ telle que :

$$\varphi(\partial_x f, \partial_y f, \partial_x^2 f, \dots). \quad (3.69)$$

Par exemple,

$$\varphi(\partial_x f, \partial_y f, \partial_x^2 f, \dots) = \partial_x^2 f + \partial_y^2 f \quad (3.70)$$

donne le Laplacien. Ici φ est linéaire et invariant par rotation.

Soit:

$$\varphi(\partial_x f, \partial_y f, \partial_x^2 f, \dots) = \sqrt{(\partial_x f)^2 + (\partial_y f)^2} \quad (3.71)$$

donne la norme du gradient. Ici φ est non linéaire mais invariant par rotation.

Soit:

$$\varphi(\partial_x f, \partial_y f, \partial_x^2 f, \dots) = \partial_{xy} f \quad (3.72)$$

correspond à φ linéaire mais dépendant de l'orientation.

Nous allons nous intéresser à la recherche d'opérateurs invariants par rotation et symétrie axiale. Dans la suite du chapitre, φ désigne un opérateur différentiel linéaire ou non (nous nous intéresserons surtout au cas polynomial).

3.5.2 Rotation de f

Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une image en niveau de gris. On note X la variable continue de \mathbb{R}^2 , $X = (x, y)$. Soit g_θ qui à le même profil que f mais avec une rotation d'angle $-\theta$ et de centre 0 :

$$g_\theta(X) = f(R_\theta X) \quad (3.73)$$

Concrètement, R_θ est la matrice 2×2 telle que :

$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \quad (3.74)$$

Soient $\delta X^1, \dots, \delta X^k$, k vecteurs de \mathbb{R}^2 , le calcul de la différentielle d'ordre k de g_θ donne:

$$d^k g_\theta(X)(\delta X^1, \dots, \delta X^k) = d^k f(R_\theta X)(R_\theta \delta X^1, \dots, R_\theta \delta X^k) \quad (3.75)$$

On s'intéresse à la différentielle en 0 :

$$d^k g_\theta(0)(\delta X^1, \dots, \delta X^k) = d^k f(0)(R_\theta \delta X^1, \dots, R_\theta \delta X^k) \quad (3.76)$$

$d^k g_\theta(0)$ et $d^k f(0)$ sont des formes k -linéaires symétriques.

Considerons maintenant la rotation d'angle $-(\theta + \delta\theta)$:

$$d^k g_{\theta+\delta\theta}(0)(\delta X^1, \dots, \delta X^k) = d^k f(0)(R_{\theta+\delta\theta} \delta X^1, \dots, R_{\theta+\delta\theta} \delta X^k) \quad (3.77)$$

Sachant que :

$$R_{\theta+\delta\theta} = R_\theta + \delta\theta R_{\theta+\frac{\pi}{2}} + o(\delta\theta) \quad (3.78)$$

On peut calculer la différentielle de $\theta \mapsto d^k g_\theta(0)$:

$$\begin{aligned} d^k g_{\theta+\delta\theta}(0)(\delta X^1, \dots, \delta X^k) &= d^k f(0)(R_\theta \delta X^1, \dots, R_\theta \delta X^k) + \\ &\delta\theta \left[\sum_{j=1}^k d^k f(0)(R_\theta \delta X^1, \dots, R_{\theta+\frac{\pi}{2}} \delta X^j, \dots, R_\theta \delta X^k) \right] + o(\delta\theta) \end{aligned} \quad (3.79)$$

De manière à pouvoir utiliser les variables x et y dans des expressions mettant en jeu des sommes et des tenseurs de manière fluide, nous allons, dans certaines expressions les remplacer par des notations indicées x_1 et x_2 :

$$x_i = \begin{cases} x & \text{si } i = 1 \\ y & \text{si } i = 2 \end{cases} \quad (3.80)$$

En tant que forme k -linéaire symétrique, $d^k f(0)$ est un tenseur de taille $\underbrace{2 \times 2 \times \dots \times 2}_{k \text{ fois}}$ de coefficients $(a_{i_1, \dots, i_k})_{\{i_1, \dots, i_k\} \in \{1, 2\}^k}$. Concrètement $d^k f(0)$ est un polynôme de degré k sur les composantes des δx^j on a :

$$d^k f(0)(\delta X^1, \dots, \delta X^k) = \sum_{i_1, \dots, i_k=1}^2 a_{i_1, \dots, i_k} \delta x_{i_1}^1 \dots \delta x_{i_k}^k \quad (3.81)$$

Ensuite, le tenseur $d^k f(0)$ est symétrique ce qui signifie sur les coefficients que :

$$\forall \sigma \in \mathfrak{S}^k, \quad a_{\sigma(i_1), \dots, \sigma(i_k)} = a_{i_1, \dots, i_k} \quad (3.82)$$

Où \mathfrak{S}^n est l'ensemble des permutations de $[[1, n]]$.

En effet, on a :

$$a_{i_1, \dots, i_k} = (\partial_{x_{i_1}} \dots \partial_{x_{i_k}} f)(0) \quad (3.83)$$

qui donne le même résultat après permutation des dérivées partielles pour une fonction C^k (théorème de Schwarz).

Finalement, seul le nombre de 1 et de 2 parmi les indices i_1, \dots, i_k influe sur la valeur de a_{i_1, \dots, i_k} .

Notons a_p^k le coefficient a_{i_1, \dots, i_k} avec $p = \text{Card}(\{i_j, i_j = 2, j \in [[1, k]]\})$ (le cardinal Card désigne le nombre d'élément dans un ensemble), cette notation va être utilisée après le paragraphe suivant. La valeur de a_p^k est alors :

$$a_p^k = \partial_x^{k-p} \partial_y^p f(0) \quad (3.84)$$

En remplaçant $R_\theta \delta X^j$ par δX^j , l'équation 3.79 fait intervenir le tenseur :

$$\sum_{j=1}^k d^k f(0)(\delta X^1, \dots, R_{\frac{\pi}{2}} \delta X^j, \dots, \delta X^k) \quad (3.85)$$

Calculons les coefficients associés à ce tenseur.

On commence avec un terme de la somme : soit $j \in [1, k]$,

$$\begin{aligned} d^k f(0)(\delta X^1, \dots, R_{\frac{\pi}{2}} \delta X^j, \dots, \delta X^k) &= \sum_{i_1, \dots, i_k=1}^2 a_{i_1, \dots, i_k} \delta x_{i_1}^1 \dots \left[R_{\frac{\pi}{2}} \delta X^j \right]_{i_j} \dots \delta x_{i_k}^k \\ &= \sum_{i_1, \dots, i_k=1}^2 a_{i_1, \dots, i_k} \delta x_{i_1}^1 \dots (-1)^{i_j} \delta x_{3-i_j}^j \dots \delta x_{i_k}^k = \sum_{i_1, \dots, i_k=1}^2 (-1)^{i_j+1} a_{i_1, \dots, 3-i_j, \dots, i_k} \delta x_{i_1}^1 \dots \delta x_{i_j}^j \dots \delta x_{i_k}^k \end{aligned} \quad (3.86)$$

Le terme en $3 - i_j$ est simplement une interversion des valeurs 1 et 2 :

$$3 - i_j = \begin{cases} 2 & \text{si } i_j = 1 \\ 1 & \text{si } i_j = 2 \end{cases} \quad (3.88)$$

Ainsi, le coefficient b_{i_1, \dots, i_k} en i_1, \dots, i_k du tenseur de l'équation 3.85 est :

$$b_{i_1, \dots, i_k} = \sum_{j=1}^k (-1)^{i_j+1} a_{i_1, \dots, 3-i_j, \dots, i_k} \quad (3.89)$$

b_{i_1, \dots, i_k} est finalement une somme de coefficients en $a_{i_1, \dots, 3-i_j, \dots, i_k}$. Nous savons que, par symétrie, seul le nombre d'indice valant 1 ou 2 influe sur la valeur de $a_{i_1, \dots, 3-i_j, \dots, i_k}$, soit $p = \text{Card}(\{i_j, i_j = 2, j \in [1, k]\})$:

$$(-1)^{i_j+1} a_{i_1, \dots, 3-i_j, \dots, i_k} = \begin{cases} a_{p+1}^k & \text{si } i_j = 1 \\ -a_{p-1}^k & \text{si } i_j = 2 \end{cases} \quad (3.90)$$

Étant donné i_1, \dots, i_k , on rencontre $(k + 1 - p)$ fois le nombre 1 et p fois le nombre 2, on a donc la valeur de b_{i_1, \dots, i_k} :

$$b_{i_1, \dots, i_k} = -p a_{p-1}^k + (k + 1 - p) a_{p+1}^k \quad (3.91)$$

Étant donné que les tenseurs $d^k f(0)$ peuvent être représentés par des "matrices" de dimension k , nous allons simplifier la notation de l'équation 3.79 :

$$d^k g_{\theta+\delta\theta}(0)(\delta X^1, \dots, \delta X^k) = A_k + \delta\theta B_k + o(\delta\theta) \quad (3.92)$$

3.5.3 Opérateur différentiel φ

Soit φ un opérateur réel :

$$(d^1 f(0), \dots, d^K f(0)) \longmapsto \varphi(d^1 f(0), \dots, d^K f(0)) \quad (3.93)$$

En poursuivant les calculs du paragraphe précédent, on obtient une condition nécessaire et suffisante sur φ pour que l'opérateur différentiel associé soit invariant par rotation :

$$\langle \nabla_a \varphi(a), Aa \rangle = 0 \quad (3.94)$$

où :

$$a = \begin{pmatrix} \partial_x f \\ \partial_y f \\ \partial_x^2 f \\ \partial_{xy} f \\ \vdots \end{pmatrix} \quad A = \begin{pmatrix} A_1 & 0 & & 0 \\ 0 & A_2 & \ddots & \\ & \ddots & & 0 \\ 0 & & & A_K \end{pmatrix} \quad A_k = \begin{pmatrix} 0 & k & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & & & -k & 0 \end{pmatrix} \quad (3.95)$$

A_k est de taille $(k+1) \times (k+1)$.

$$a_i^k = \partial_x^{k-i} \partial_y^i f(0,0) \quad i \in [0, k] \quad (3.96)$$

où k est l'ordre de la dérivée partielle, i est l'ordre de la dérivée selon y .

$$a^k = \begin{pmatrix} a_0^k \\ a_1^k \\ \vdots \\ a_i^k \\ \vdots \\ a_k^k \end{pmatrix} = \begin{pmatrix} \partial_x^k f(0,0) \\ \partial_x^{k-1} \partial_y f(0,0) \\ \vdots \\ \partial_x^{k-i} \partial_y^i f(0,0) \\ \vdots \\ \partial_y^k f(0,0) \end{pmatrix} \quad (3.97)$$

Dans les parties qui suivent, nous allons chercher des polynômes de degré n de dérivées partielles invariants par rotation. Par exemple si $n = 1$, l'opérateur φ est linéaire, si $n = 2$, il est quadratique, etc. k désigne l'ordre des dérivées partielles, on nommera k_1, k_2 etc si différents ordres interviennent, lorsqu'il n'y a qu'un seul ordre nous dirons que φ est homogène.

3.5.4 Recherche des opérateurs différentiels invariants par rotation et linéaires (n=1)

Soit φ une application linéaire sur les dérivées a :

$$\varphi(a) = \sum_{k=1}^K \sum_{i=0}^k \lambda_i^k a_i^k \quad (3.98)$$

λ_i^k est le coefficient associé à la dérivée a_i^k .

Notons :

$$\lambda^k = \begin{pmatrix} \lambda_0^k \\ \vdots \\ \lambda_k^k \end{pmatrix} \quad \text{et} \quad \lambda = \begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda^K \end{pmatrix} = \begin{pmatrix} \lambda_0^1 \\ \lambda_1^1 \\ \lambda_0^2 \\ \vdots \\ \lambda_K^K \end{pmatrix} \quad \text{on a alors :} \quad \varphi(a) = \lambda \cdot a \quad (3.99)$$

On a :

$$\nabla_a \varphi(a) = \lambda \quad (3.100)$$

Donc :

$$\langle \nabla_a \varphi(a), Aa \rangle = \langle A^T \lambda, a \rangle = \sum_{k=1}^K \langle (A_k)^T \lambda^k, a^k \rangle \quad (3.101)$$

$$\partial_\theta \varphi(a) = 0 \Leftrightarrow A^T \lambda = 0 \Leftrightarrow \forall k \in [1, K], (A_k)^T \lambda^k = 0 \quad (3.102)$$

On a alors que tout opérateur linéaire invariant par rotation peut être décomposé en opérateurs linéaires invariants par rotation faisant intervenir des dérivées du même ordre.

Il n'y a pas d'application linéaire non nulle invariante par rotation faisant intervenir des dérivées d'ordre 1 :

$$A_1^T \lambda^1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_0^1 \\ \lambda_1^1 \end{pmatrix} = 0 \Leftrightarrow \lambda_0^1 = 0 \text{ et } \lambda_1^1 = 0 \quad (3.103)$$

Plus généralement, lorsque $\det(A_k^T) \neq 0$, alors il n'y a pas d'application linéaire non nulle invariante par rotation faisant intervenir des dérivées d'ordre k .

Les matrices A_k^T ont des coefficients non nuls sur les sous diagonales et nuls partout ailleurs

$((A_k^T)[1, j] \neq 0 \Leftrightarrow |i - j| = 1)$.

Soit B_n une matrice de cette forme $((B_n)_{i,j} \neq 0 \Leftrightarrow |i - j| = 1)$ de taille $n \times n$ avec $n \geq 3$.

$$\det(B_n) = \begin{vmatrix} 0 & b & 0 & & \cdots & 0 \\ a & 0 & d & 0 & \cdots & 0 \\ 0 & c & & & & \\ & 0 & & B_{n-2} & & \\ \vdots & \vdots & & & & \\ 0 & 0 & & & & \end{vmatrix} = -a \begin{vmatrix} b & 0 & \cdots & 0 \\ c & & & \\ 0 & & & \\ \vdots & & & B_{n-2} \\ 0 & & & \end{vmatrix} = -ab \det(B_{n-2}) \quad (3.104)$$

Comme $a \neq 0$ et $b \neq 0$, on a :

$$\det(B_n) \neq 0 \Leftrightarrow \det(B_{n-2}) \neq 0 \quad (3.105)$$

Comme on a, pour $a, b, c, d \neq 0$:

$$\begin{vmatrix} 0 & b \\ a & 0 \end{vmatrix} = -ab \neq 0 \quad \text{et} \quad \begin{vmatrix} 0 & b & 0 \\ a & 0 & d \\ 0 & c & 0 \end{vmatrix} = 0 \quad (3.106)$$

Alors :

$$\det(A_k^T) \neq 0 \text{ pour } k \text{ impair} \quad \text{et} \quad \det(A_k^T) = 0 \text{ pour } k \text{ pair} \quad (3.107)$$

A_k^T est inversible pour k impair, donc $A_k^T \lambda^k = 0 \Rightarrow \lambda^k = 0$.

Donc il n'y a pas d'application linéaire non nulle invariante par rotation faisant intervenir des dérivées d'ordre impair, mais il y en a toujours au moins une pour les ordres pairs.

Pour $k = 2p$, on peut séparer le système $A_{2p}^T \lambda^{2p} = 0$ en deux parties :

$$\begin{cases} -\lambda_1^{2p} & = 0 \\ (2p+1)\lambda_3^{2p} - 3\lambda_1^{2p} & = 0 \\ \cdots & \\ (2p+1-2j)\lambda_{2j+1}^{2p} - (2j+1)\lambda_{2j-1}^{2p} & = 0 \end{cases} \quad (3.108)$$

$$\begin{cases} (2p)\lambda_0^{2p} & = 0 \\ (2p-2)\lambda_2^{2p} - 4\lambda_0^{2p} & = 0 \\ \cdots & \\ (2p-2j)\lambda_{2j}^{2p} - (2j+2)\lambda_{2j+2}^{2p} & = 0 \end{cases} \quad (3.109)$$

Le système (3.108) montre que les coefficients λ_i^{2p} avec i impair, sont nuls :

$$\begin{aligned} \lambda_1^{2p} &= 0 \\ (2p+1)\lambda_3^{2p} - 3\lambda_1^{2p} &= 0 \Rightarrow \lambda_3^{2p} = 0 \\ \cdots & \end{aligned} \quad (3.110)$$

par récurrence $\lambda_{2j-1}^{2p} = 0$ et $(2p+1-2j)\lambda_{2j+1}^{2p} - (2j+1)\lambda_{2j-1}^{2p} = 0 \Rightarrow \lambda_{2j+1}^{2p} = 0$

A propos du système (3.109), on a :

$$2p\lambda_0^{2p} - 2\lambda_2^{2p} = 0 \Rightarrow \lambda_2^{2p} = p\lambda_0^{2p} \quad (3.111)$$

$$(2p-2)\lambda_2^{2p} - 4\lambda_4^{2p} = 0 \Rightarrow \lambda_4^{2p} = \frac{p-1}{2}\lambda_2^{2p} = \frac{p(p-1)}{2}\lambda_0^{2p} = \binom{p}{2}\lambda_0^{2p} \quad (3.112)$$

La notation $\binom{p}{j}$ désigne le coefficient binomial $\frac{p!}{j!(p-j)!}$.

Par récurrence, supposons que $\lambda_{2j}^{2k} = \binom{p}{j}\lambda_0^{2k}$

$$(2p-2j)\lambda_{2j}^{2p} - (2j+2)\lambda_{2j+2}^{2p} = 0 \Rightarrow \lambda_{2j+2}^{2p} = \frac{p-j}{j+1}\binom{p}{j}\lambda_0^{2p} \quad (3.113)$$

or :

$$\frac{p-j}{j+1} \binom{p}{j} = \frac{(p-j) \cdot p!}{(j+1) \cdot j! \cdot (p-j)!} = \frac{p!}{(j+1)! \cdot (p-(j+1))!} = \binom{p}{j+1} \quad (3.114)$$

Donc l'espace des solutions de $A_{2p}^T \lambda^{2p} = 0$ est une droite vectorielle, il n'y a qu'une application linéaire invariante par rotation.

En posant $\lambda_0^{2p} = 1$, on a : $\lambda_{2j}^{2k} = \binom{p}{j}$

$$\varphi(a^{2p}) = \sum_{j=0}^p \binom{p}{j} a_{2j}^{2p} = \sum_{j=0}^p \binom{p}{j} \partial_x^{2p-2j} \partial_y^{2j} f = (\partial_x^2 + \partial_y^2)^p f = \Delta^p f \quad (3.115)$$

En ajoutant le cas $k = 0$ qui correspond à une dérivées d'ordre 0, c'est-à-dire l'identité f , nous avons le tableau récapitulatif des opérateurs différentiels linéaires invariants par rotation :

ordre de la dérivée k	0	1	2	3	4	5	6	7	...	$2p$	$2p+1$...
opérateur différentiel	f	-	Δf	-	$\Delta^2 f$	-	$\Delta^3 f$	-	...	$\Delta^p f$	-	...

3.5.5 Recherche des opérateurs différentiels invariants par rotation quadratiques et homogènes ($n = 2$, $k_1 = k_2$)

Nous pouvons commencer par chercher des opérateurs différentiels s'exprimant comme polynôme de degré 2 de dérivées partielles de même ordre.

$$\varphi(a^k) = \frac{1}{2} M : a^k \otimes a^k = \frac{1}{2} (a^k)^T M a^k \quad (3.116)$$

où $M = (M_{i,j})_{i,j \in [0,k]}$ est une matrice symétrique $(k+1) \times (k+1)$. Le $\frac{1}{2}$, de même que le $\frac{1}{n}$ plus tard, est introduit pour simplifier les notation dans la suite.

Les valeurs des coefficients de M permettront de trouver les valeurs des coefficients devant les dérivées partielles dans l'expression de l'opérateur.

Par exemple pour $k = 3$, le coefficient $M_{0,2}$ sera associé au produit de dérivées partielles $\partial_x^3 f \cdot \partial_x \partial_y^2 f$, ce coefficient sera $2M_{0,2}$.

$$\nabla_a \varphi(a) = M a^k \quad (3.117)$$

Par exemple, pour $k = 1$ (cas simple), nous avons :

$$\langle \nabla_a \varphi(a), A_1 a \rangle = a^T A_1^T M a \quad \text{avec} \quad A_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (3.118)$$

(ici a désigne a^1)

Les inconnus sont les coefficients de M , sachant que $M_{0,1} = M_{1,0}$, il y a trois inconnues : $m_1 = M_{0,0}$, $m_2 = M_{0,1}$ et $m_3 = M_{1,1}$. (Renommer ainsi les coefficients de M sera plus pertinent pour k et/ou n grand et pour l'implémentation.)

$$a^T A_1^T M a = -m_2 (a_0)^2 + (m_1 - m_3) a_0 a_1 + m_2 (a_1)^2 \quad (3.119)$$

Les coefficients doivent alors vérifier :

$$\begin{cases} -m_2 & = & 0 \\ m_1 - m_3 & = & 0 \\ m_2 & = & 0 \end{cases} \Rightarrow \begin{cases} m_3 & = & m_1 \\ m_2 & = & 0 \end{cases} \quad (3.120)$$

Le noyau de l'application linéaire $M \mapsto \langle \nabla_a \varphi(a), A_1 a \rangle$ est de dimension 1, il y a donc un seul opérateur différentiel correspondant à la matrice $M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

$$a^T M a = (a_0)^2 + (a_1)^2 = (\partial_x f)^2 + (\partial_y f)^2 = \|\nabla f\|^2 \quad (3.121)$$

Pour $k \in \mathbb{N}^*$

M est une matrice $(k+1) \times (k+1)$ symétrique. Soit T l'ensemble des représentants des classes d'équivalence à permutation près d'indices. Et ψ une fonction qui numérote ces classes d'indice.

$$T = \{(i, j), 0 \leq i \leq j \leq k\}, \quad \psi : T \rightarrow \llbracket 1, \text{Card}(T) \rrbracket \quad (3.122)$$

Ici, $\text{Card}(T) = \frac{(k+1)(k+2)}{2} = \binom{k+2}{2}$. On peut choisir l'ordre lexicographique pour le comptage de T par ψ , c'est-à-dire : $(0, 0), (0, 1), (0, 2), \dots (0, k), (1, 1), (1, 2), \dots (k-1, k), (k, k)$. On étend ψ sur chaque classe d'équivalence.

$$\varphi(a) = \frac{1}{2} M : a \otimes a \quad (3.123)$$

$$\partial_\theta \varphi(a) = \langle A_k a, (M : e_{i_1} \otimes a)_{i_1} \rangle = \sum_{i_1, i_2, i_3=0}^k (A_k)_{i_1, i_3} M_{i_1, i_2} a_{i_2} a_{i_3} \quad (3.124)$$

(ici a désigne a^k)

Le polynôme du second degré $\partial_\theta \varphi(a)$ est nul, ce qui signifie que la somme des coefficients correspondant à un même $a_{i_2} a_{i_3}$ sont nuls.

On en déduit un système d'inconnu le vecteur $(m_i)_{i \in \llbracket 1, \text{Card}(T) \rrbracket}$:

$$\text{Ligne } I = \psi(i_2, i_3) : \quad \sum_{i_1, i_2, i_3 \in \llbracket 0, k \rrbracket, \psi(i_2, i_3) = I} (A_k)_{i_1, i_3} m_{\psi(i_1, i_2)} = 0 \quad (3.125)$$

Pour $k = 2$,

$(i_1, i_2) \in T$	$\psi(i_1, i_2)$
(0, 0)	1
(0, 1)	2
(0, 2)	3
(1, 1)	4
(1, 2)	5
(2, 2)	6

On peut établir le système d'inconnues de plusieurs manières (par exemple en développant $a^T A_2^T M a$, en disant que $A_2^T M$ est antisymétrique, ou en établissant les equations de chaque ligne avec la formule (3.125)).

Le système obtenu est :

$$\begin{cases} -m_2 & = & 0 \\ 2m_1 - 2m_3 - m_4 & = & 0 \\ m_2 - m_5 & = & 0 \\ 2m_2 - 2m_5 & = & 0 \\ 2m_3 + m_4 - 2m_6 & = & 0 \\ m_5 & = & 0 \end{cases} \implies \begin{cases} m_2 & = & 0 \\ m_6 & = & m_1 \\ 2m_3 & = & -m_4 + 2m_1 \\ m_5 & = & 0 \end{cases} \quad (3.126)$$

Si on pose :

$$m_1 = u \quad \text{et} \quad m_4 = 2v \quad \text{on obtient :} \quad \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \end{pmatrix} = u \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} + v \begin{pmatrix} 0 \\ 0 \\ -1 \\ 2 \\ 0 \\ 0 \end{pmatrix} \quad (3.127)$$

C'est une base du noyau de l'application linéaire sur m dont on a une expression avec le système. Ce qui correspond à deux matrices :

$$M_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{et} \quad M_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 0 \end{pmatrix} \quad (3.128)$$

$$a^T M_1 a = (a_0)^2 + 2a_0 a_2 + (a_2)^2 = (\partial_x^2 f)^2 + 2\partial_x^2 f \partial_y^2 f + (\partial_y^2 f)^2 = (\Delta f)^2 \quad (3.129)$$

$$a^T M_2 a = -2a_0 a_2 + 2(a_1)^2 = -2\partial_x^2 f \partial_y^2 f + 2(\partial_x \partial_y f)^2 = -2 \det(H_f) \quad (3.130)$$

où H_f est la matrice hessienne de f .

Donc dans le cas quadratique, pour des dérivées partielles d'ordre 2, nous avons deux opérateurs : $(\Delta f)^2$ et $\det(H_f)$.

Voici un tableau montrant les opérateurs non linéaires pour un degré inférieur ou égal à 4 invariants par rotation:

k	$\varphi(a)$
0	f^2
1	$\ \nabla f\ ^2$
2	$(\Delta f)^2$ $\det(H_f)$
3	$(\partial_x^3 f)^2 + 3\partial_x^3 f \cdot \partial_x \partial_y^2 f + 3\partial_x^2 \partial_y f \cdot \partial_y^3 f + (\partial_y^3 f)^2$ $(\partial_x^3 f)^2 + 3(\partial_x^2 \partial_y f)^2 + 3(\partial_x \partial_y^2 f)^2 + (\partial_y^3 f)^2$
4	$(\Delta^2 f)^2$ $\det(H_{\Delta f})$ $\partial_x^4 f \cdot \partial_y^4 f - 4\partial_x^3 \partial_y f \cdot \partial_x \partial_y^3 f + 3(\partial_x^2 \partial_y^2 f)^2$

3.5.6 Recherche des opérateurs quadratiques invariants par rotation et non homogènes ($n = 2$, $k_1 \neq k_2$)

On début de la section précédente, nous avons dit que nous commençons par des polynômes faisant intervenir des dérivées partielles de même degré. Il existe aussi des polynômes tels que : $f \cdot \Delta f$, $f \cdot \Delta^2 f$ ou $\Delta f \cdot \Delta^2 f$ qui sont invariants par rotation.

Si on cherche des opérateurs de degré 2 faisant intervenir des dérivées d'ordre 2 et 1, on a une matrice M qui est de taille 3×2 :

$$\varphi(a) = M : a^2 \otimes a^1 \quad (3.131)$$

$$\partial_\theta \varphi(a) = \langle A_2 a^2, (M : e_{i_1} \otimes a^1)_{i_1 \in \llbracket 0,2 \rrbracket} \rangle + \langle A_1 a^1, (M : a^2 \otimes e_{i_1})_{i_1 \in \llbracket 0,1 \rrbracket} \rangle \quad (3.132)$$

Cette fois-ci, M n'est pas symétrique (elle n'est même plus carrée).

Le système obtenu sur les coefficients de M est :

$$\begin{cases} -M_{1,0} - M_{0,1} & = 0 \\ 2M_{0,0} - 2M_{2,0} - M_{1,1} & = 0 \\ M_{1,0} - M_{2,1} & = 0 \\ M_{0,0} - M_{1,1} & = 0 \\ 2M_{0,1} - 2M_{2,1} + M_{1,0} & = 0 \\ M_{1,1} + M_{2,0} & = 0 \end{cases} \quad (3.133)$$

Le système nous donne que tous les coefficients sont nuls, il n'y a donc pas d'opérateurs invariant par rotation de degré 2 avec des dérivées partielles d'ordre 1 et 2.

Dans le cas du degré 2 avec deux ordres k_1 et k_2 distincts, on a :

$$\varphi(a) = M : a^{k_1} \otimes a^{k_2} \quad (3.134)$$

Avec M de taille $(k_1 + 1) \times (k_2 + 1)$.

$$\partial_\theta \varphi(a) = \langle A_{k_1} a^{k_1}, (M : e_{i_1} \otimes a^{k_2})_{i_1} \rangle + \langle A_{k_2} a^{k_2}, (M : a^{k_1} \otimes e_{i_1})_{i_1} \rangle \quad (3.135)$$

$$= \sum_{i_1=0}^{k_1} \sum_{i_2=0}^{k_2} \sum_{i_3=0}^{k_1} (A_{k_1})_{i_1, i_3} M_{i_1, i_2} a_{i_3}^{k_1} a_{i_2}^{k_2} + \sum_{i_1=0}^{k_2} \sum_{i_2=0}^{k_1} \sum_{i_3=0}^{k_2} (A_{k_2})_{i_1, i_3} M_{i_2, i_1} a_{i_2}^{k_1} a_{i_3}^{k_2} \quad (3.136)$$

Dans les deux sommes faisant intervenir i_3 , on peut simplifier par $\sum_{i_3=\max(0, i_1-1)}^{\min(k_1, i_1+1)}$ et $\sum_{i_3=\max(0, i_1-1)}^{\min(k_2, i_1+1)}$, car $(A_k)_{i_1, i_3} \neq 0 \Rightarrow |i_1 - i_3| \leq 1$.

Ici les indices sont dans $[[0, k_1]] \times [[0, k_2]]$ (pas de symétrie ici car les produits $a_i^{k_1} a_j^{k_2}$ sont différents pour des couples (i, j) différents), soit ψ la fonction de comptage :

$$\psi(i, j) = i(k_2 + 1) + j + 1 \quad (3.137)$$

Algorithmiquement, on a deux ensembles de trois boucles for :

On crée une matrice nulle B de taille $(k_1 + 1)(k_2 + 1) \times (k_1 + 1)(k_2 + 1)$.

somme 1 :

Pour $i_1 = 0 : k_1, i_2 = 0 : k_2, i_3 = \max(0, i_1 - 1), \min(k_1, i_1 + 1) :$

$$I = \psi(i_3, i_2), J = \psi(i_1, i_2), B[I, J] = B[I, J] + (A_{k_1})_{i_1, i_3} \quad (3.138)$$

somme 2 :

Pour $i_1 = 0 : k_2, i_2 = 0 : k_1, i_3 = \max(0, i_1 - 1), \min(k_2, i_1 + 1) :$

$$I = \psi(i_2, i_3), J = \psi(i_2, i_1), B[I, J] = B[I, J] + (A_{k_2})_{i_1, i_3} \quad (3.139)$$

Voici, dans le tableau, les opérateurs trouvés suivant les valeurs de k_1 et k_2 (la diagonale, avec les points de suspension, correspond au tableau déjà établi pour le cas homogène).

(k_1, k_2)	0	1	2	3	4	5
0	...	-	$f \cdot \Delta f$	-	$f \cdot \Delta^2 f$	-
1		...	-	$\langle \nabla f, \nabla \Delta f \rangle$ $\det(\nabla f, \nabla \Delta f)$	-	$\langle \nabla f, \nabla \Delta^2 f \rangle$ $\det(\nabla f, \nabla \Delta^2 f)$
2			...	-	$(\partial_x^2 f \cdot \partial_y^2 - 2\partial_x \partial_y f \cdot \partial_x \partial_y + \partial_y^2 f \cdot \partial_x^2) \Delta f$ $((-\partial_x^2 + \partial_y^2) f \cdot \partial_x \partial_y + \partial_x \partial_y f \cdot (\partial_x^2 - \partial_y^2)) \Delta f$ $((\partial_x^2 f - \partial_y^2 f) \cdot (\partial_x^2 - \partial_y^2) + 4\partial_x \partial_y f \cdot \partial_x \partial_y) \Delta f$	-
3				...	-	(4 opérateurs)
4					...	-
5						...

Ce n'est pas toujours facile de trouver une expression condensée pour ces opérateurs, il est possible que le cas $k_1 = 2, k_2 = 4$ ait des expressions réduites de ces opérateurs.

Une utilisation formelle du symbole ∇ peut simplifier les expressions en :

$$((H_f - \Delta f I_2) : \nabla \otimes \nabla)(\Delta f), \quad (H_f : \nabla \otimes \nabla)(\Delta f), \quad (R_{\frac{\pi}{2}} H_f : \nabla \otimes \nabla)(\Delta f) \quad (3.140)$$

On peut utiliser la notation suivante mais potentiellement ambiguë :

$$(\nabla^T \cdot H_f \cdot \nabla)(\Delta f) \quad (3.141)$$

mais par contre cet opérateur est différent de :

$$(\nabla(\Delta f))^T \cdot H_f \cdot (\nabla(\Delta f)) \quad (3.142)$$

qui lui est de degré 3.

3.5.7 Recherche des opérateurs différentiels invariants par rotation, de degré n et homogènes

$$\varphi(a) = \frac{1}{n} M :_n a^k \otimes \dots \otimes a^k \quad (3.143)$$

M est un tenseur symétrique de dimension n de taille $(k+1) \times \dots \times (k+1)$.

$$\partial_\theta \varphi(a) = \langle A_k a^k, (M :_n e_{i_1} \otimes a^k \otimes \dots \otimes a^k)_{i_1} \rangle \quad (3.144)$$

$$= \sum_{i_1=0}^k (A_k a^k)[i_1] \cdot M :_n e_{i_1} \otimes a^k \otimes \dots \otimes a^k \quad (3.145)$$

$$M :_n e_{i_1} \otimes a^k \otimes \dots \otimes a^k = \sum_{i_2, \dots, i_n=0}^k M_{i_1, i_2, \dots, i_n} a_{i_2}^k \cdot a_{i_3}^k \dots a_{i_n}^k \quad (3.146)$$

$$(A_k a^k)[i_1] = \sum_{i_{n+1}=0}^k (A_k)_{i_1, i_{n+1}} a_{i_{n+1}}^k \quad (3.147)$$

Finalement nous cherchons les valeurs des coefficients de M de manière à ce que le polynôme de degré n en a^k soit nul :

$$\sum_{i_1, i_2, \dots, i_n, i_{n+1}} (A_k)_{i_1, i_{n+1}} M_{i_1, i_2, \dots, i_n} a_{i_2}^k \cdot a_{i_3}^k \dots a_{i_n}^k a_{i_{n+1}}^k = 0 \quad (3.148)$$

Soit T , l'ensemble de représentants de classe d'équivalence d'indices à permutation près :

$$T = \{(j_1, \dots, j_n), 0 \leq j_1 \leq j_2 \leq \dots \leq j_n \leq k\} \quad (3.149)$$

Le cardinal de T est de $C_{n+k}^n = \binom{k+n}{n}$ car l'application $(i_1, i_2, \dots, i_n) \mapsto (i_1+1, i_2+2, \dots, i_n+n)$ est bijective de T dans l'ensemble des parties à n éléments de $[[1, k+n]]$.

Soit ψ bijective qui compte les éléments de T , par exemple dans l'ordre lexicographique :

$$\psi : T \rightarrow [[1, \text{Card}(T)]] \quad (3.150)$$

On étend ψ sur les classes d'équivalence, elle est alors invariante par permutation des indices.

$$\text{Ligne } I = \psi(i_2, i_3, \dots, i_{n+1}) : \sum_{\substack{i_1, i_2, \dots, i_{n+1} \in [[0, k]], \\ \psi(i_2, i_3, \dots, i_{n+1}) = I}} (A_k)_{i_1, i_{n+1}} m_{\psi(i_1, i_2, \dots, i_n)} = 0 \quad (3.151)$$

Le calcul à la main devient très vite laborieux, mais par résolution numérique, on peut établir la matrice du système :

On crée un tableau B de taille $C_{n+k}^n \times C_{n+k}^n$ avec des valeurs nulles, on parcourt les indices i_1, i_2, \dots, i_{n+1} , on pose $I = \psi(i_2, i_3, \dots, i_{n+1})$ et $J = \psi(i_1, i_2, \dots, i_n)$ et on remplit le tableau :

$$B[I, J] = B[I, J] + (A_k)[i_1, i_{n+1}] \quad (3.152)$$

Etant donné que les coefficients de K_k sont entiers, B est aussi à coefficients entiers. Les coefficients des polynômes qu'on recherche seront donc rationnels.

Souvent les logiciels de calculs ont une fonction qui donne le rang d'une matrice, on peut en déduire la dimension du noyau de la matrice grâce au théorème du rang :

Théorème du rang : Si f est une application linéaire entre deux espaces vectoriels E et F ($f \in \mathcal{L}(E, F)$), alors :

$$\text{rang}(f) + \dim(\text{Ker}(f)) = \dim(E) \quad (3.153)$$

Donc, $C_{n+k}^n - \text{rang}(B)$ nous donne la dimension du noyau de B , ce qui correspond au nombre maximal de polynômes de degré n de dérivées partielles d'ordre k invariants par rotation linéairement indépendants.

Matlab propose la fonction *null* qui renvoie une base du noyau de la matrice en argument, incluant des options de bases normées ou avec des coefficients rationnels.

Les coefficients de cette base du noyau de B ne sont pas encore les coefficients des polynômes car il faut les multiplier par le nombre d'anagrammes des indices i_1, \dots, i_n associé à J . Soit c_j le nombre parmi les indices i_1, \dots, i_n valant j . Le nombre d'anagrammes est alors :

$$\frac{n!}{c_0!c_1!\dots c_k!} \quad (3.154)$$

Par exemple, pour $k = 2$ et $n = 3$, $J = 3$ correspond aux indices $(i_1, i_2, i_3) = (0, 0, 2)$ dans T mais il correspond à trois coefficients de M : $M_{0,0,2}$, $M_{0,2,0}$ et $M_{2,0,0}$ qui sont égaux, mais dans le calcul de $\phi(a)$, le coefficient associé à $(a_0^2)^2 a_2^2$ est $M(0, 0, 2) + M(0, 2, 0) + M(2, 0, 0) = 3m_3$, ce coefficient 3 devant m_3 est le nombre d'anagrammes de $(0, 0, 2)$.

3.5.7.1 Symétries axiales

Le système vérifié par le vecteur m peut être séparé en deux sous-systèmes.

Dans la formule (3.151) nous pouvons voir une correspondance entre la ligne $I = \psi(i_2, i_3, \dots, i_{n+1})$ et les coefficients $m_{\psi(i_1, i_2, \dots, i_n)}$. Etant donné que $A_k[i_1, i_{n+1}]$ est nul lorsque $|i_1 - i_{n+1}| \neq 1$, alors i_1 et i_{n+1} n'ont pas la même parité, or ce sont exactement les deux indices qui diffèrent dans l'expression (dans les arguments dans ψ) de la ligne I et de l'indice de m . Donc sur une même ligne, il n'y aura que des coefficients m dont les indices associés sont de somme $i_1 + i_2 + \dots + i_n$ paire, ou de somme impaire.

Dans les exemples de la partie quadratique, les coefficients m de somme impaire correspondent à m_2 dans le cas $k = 1$, et à m_2 et m_5 dans le cas $k = 2$. Et les λ_{2j+1}^{2p} dans la partie linéaire.

$$a_p^k = \partial_x^{k-p} \partial_y^p f(0, 0) = d^k f(0, 0) \left(\binom{1}{0}, \dots, \binom{1}{0}, \underbrace{\binom{0}{1}, \dots, \binom{0}{1}}_{p \text{ fois}} \right) \quad (3.155)$$

La même dérivée partielle mais dans les directions x et $-y$ vaut :

$$\partial_x^{k-p} \partial_{-y}^p f(0, 0) = d^k f(0, 0) \left(\binom{1}{0}, \dots, \binom{1}{0}, \underbrace{\binom{0}{-1}, \dots, \binom{0}{-1}}_{p \text{ fois}} \right) = (-1)^p a_p^k \quad (3.156)$$

Soit :

$$\varphi(a) = M :_n a^k \otimes \dots \otimes a^k \quad (3.157)$$

un polynôme de degré n de dérivées partielles d'ordre k , et :

$$S(a^k) = \begin{pmatrix} (-1)^0 a_0^k \\ (-1)^1 a_1^k \\ \vdots \\ (-1)^k a_k^k \end{pmatrix} = ((-1)^p a_p^k)_{p \in [0, k]} \quad (3.158)$$

la symétrie d'axe Ox des dérivées partielles d'ordre k .

Si φ est invariante par symétrie, alors elle vérifie :

$$M :_n S(a^k) \otimes \dots \otimes S(a^k) = M :_n a^k \otimes \dots \otimes a^k \quad (3.159)$$

Pour chaque ensemble d'indices i_1, i_2, \dots, i_n à permutation près et son nombre d'anagrammes c ($c \geq 1$) on a :

$$c M_{i_1, i_2, \dots, i_n} (-1)^{i_1 + \dots + i_n} a_{i_1}^k \dots a_{i_n}^k = c M_{i_1, i_2, \dots, i_n} a_{i_1}^k \dots a_{i_n}^k \quad (3.160)$$

Donc si φ est invariant par symétrie axiale alors ces coefficients M_{i_1, \dots, i_n} non nuls sont de somme $i_1 + \dots + i_n$ paire.

Nous avons vu en début de paragraphe que les opérateurs invariants par rotation peuvent être séparés en deux familles : ceux qui font intervenir des coefficients de somme d'indices paire, et ceux de somme d'indices impaire.

Ceux de somme d'indices impaire présente une antisymétrie axiale :

$$\varphi(S(a^k)) = -\varphi(a^k) \quad (3.161)$$

Pour en revenir aux images de coupes histologiques, les opérateurs antisymétriques ne seraient a priori pas utiles étant donné qu'une lame de tissu peut-être déposée en recto ou verso sur la lame de verre (ce qui correspond à une symétrie axiale sur l'image).

Par exemple, pour les polynômes de degré 3 on a le début de tableau suivant :

k	$n = 3$	$n = 4$
0	f^3	f^4
1	-	$\ \nabla f\ ^4$
2	$(\Delta f)^3$ $(\partial_x^2 f)^3 + 3\Delta f \cdot (\partial_x \partial_y f)^2 + (\partial_y^2 f)^3$	$(\Delta f)^4$ $(\Delta f)^2 \cdot \det(H_f)$ $(\det(H_f))^2$
3	-	(5 opérateurs)
4	(5 opérateurs) (dont 1 antisymétrique)	(8 opérateurs)
5	-	(12 opérateurs)
6	(8 opérateurs) (dont 2 antisymétriques)	(18 opérateurs)

3.5.8 Un exemple en degré 3 non homogène

On cherche les opérateurs de degré 3 dont les produits font intervenir deux dérivées partielles d'ordre 1 et une d'ordre 2.

$$\varphi(a) = M :_3 a^1 \otimes a^1 \otimes a^2 \quad (3.162)$$

M est de taille $2 \times 2 \times 3$ avec une symétrie sur seulement les deux premiers indices : $M_{i_1, i_2, i_3} = M_{i_2, i_1, i_3}$.

$$\partial_\theta \varphi(a) = 2 \langle A_1 a^1, (M :_3 e_{i_1} \otimes a^1 \otimes a^2)_{i_1} \rangle + \langle A_2 a^2, (M :_3 a^1 \otimes a^1 \otimes e_{i_1})_{i_1} \rangle \quad (3.163)$$

$$= 2 \sum_{i_1=0}^1 \sum_{i_2=0}^1 \sum_{i_3=0}^2 \sum_{i_4=0}^1 (A_1)_{i_1, i_4} M_{i_1, i_2, i_3} a_{i_2}^1 a_{i_4}^1 a_{i_3}^2 + \sum_{i_1=0}^2 \sum_{i_2=0}^1 \sum_{i_3=0}^1 \sum_{i_4=0}^2 (A_2)_{i_1, i_4} M_{i_2, i_3, i_1} a_{i_2}^1 a_{i_3}^1 a_{i_4}^2 \quad (3.164)$$

$$T = \{(i_1, i_2, i_3), 0 \leq i_1 \leq i_2 \leq 1, 0 \leq i_3 \leq 2\} \quad (3.165)$$

Ici $\text{Card}(T) = 9$.

On trouve trois opérateurs :

(1, 1, 2)	$(\nabla f)^T \cdot (H_f - \Delta f \cdot I_2) \cdot \nabla f$ $(\nabla f)^T \cdot H_f \cdot \nabla f$ $(\nabla f)^T \cdot R_{\frac{\pi}{2}} H_f \cdot \nabla f$ (antisymétrique)	où $R_{\frac{\pi}{2}}$ est la matrice de rotation d'angle $\frac{\pi}{2}$
-----------	---	---

On peut remarquer que $(\nabla f)^T \cdot (H_f - \Delta f \cdot I_2) \cdot \nabla f$ est le numérateur de l'expression développée de la courbure des lignes de même couleur sur un canal:

$$\text{div} \left(\frac{\nabla f}{\|\nabla f\|} \right) \quad (3.166)$$

(le dénominateur est $\|\nabla f\|^3$).

La quantité :

$$(\nabla f)^T \cdot (H_f - \Delta f \cdot I_2) \cdot \nabla f - \alpha \|\nabla f\|^3 \quad (3.167)$$

est positive sur des parties de l'image dont la courbure est supérieure à α .

Donc, même si la courbure est une fraction, un algorithme cherchant une application linéaire sur des observations comprenant les quantités $(\nabla f)^T \cdot (H_f - \Delta f \cdot I_2) \cdot \nabla f$ et $\|\nabla f\|^3$ pourra prendre en compte la courbure des lignes de même couleur.

Le cas (1, 1, 4) donne des résultats similaires :

(1, 1, 1)	-
(1, 1, 2)	$(\nabla f)^T \cdot (H_f - \Delta f \cdot I_2) \cdot \nabla f$ $(\nabla f)^T \cdot H_f \cdot \nabla f$ $(\nabla f)^T \cdot R_{\frac{\pi}{2}} H_f \cdot \nabla f$ (antisymétrique)
(1, 1, 3)	-
(1, 1, 4)	$(\nabla f)^T \cdot (H_{\Delta f} - \Delta^2 f \cdot I_2) \cdot \nabla f$ $(\nabla f)^T \cdot H_{\Delta f} \cdot \nabla f$ $(\nabla f)^T \cdot R_{\frac{\pi}{2}} H_{\Delta f} \cdot \nabla f$ (antisymétrique)
(1, 1, 5)	-
(1, 1, 6)	3 opérateurs
(1, 1, 7)	-

3.5.9 Opérateurs invariants sur deux fonctions

Jusqu'à présent, nous nous sommes intéressés aux opérateurs invariants sur une fonction f , mais en fait, une image en couleur est la donnée de trois canaux. On pourrait également rechercher des opérateurs faisant intervenir deux fonction ou plus.

Voici un exemple simple avec deux fonctions f et g et la recherche d'opérateurs de degré 2 faisant intervenir (dans leurs produits) une dérivée d'ordre 1 de f et une dérivée d'ordre 1 de g :

$$\varphi(a(f), a(g)) = M : a^1(f) \otimes a^1(g) \quad (3.168)$$

$$\partial_\theta \varphi(a(f), a(g)) = \langle A_1 a^1(f), (M : e_{i_1} \otimes a^1(g))_{i_1} \rangle + \langle A_1 a^1(g), (M : a^1(f) \otimes e_{i_1})_{i_1} \rangle \quad (3.169)$$

$$= \sum_{0 \leq i_1, i_2, i_3 \leq 1} A_{1i_1, i_3} M_{i_1, i_2} a_{i_3}^1(f) a_{i_2}^1(g) + \sum_{0 \leq i_1, i_2, i_3 \leq 1} A_{i_1, i_3} M_{1i_2, i_1} a_{i_2}^1(f) a_{i_3}^1(g) \quad (3.170)$$

Ce qui donne le système :

$$\begin{cases} -M_{1,0} - M_{0,1} = 0 \\ -M_{1,1} + M_{0,0} = 0 \\ M_{0,0} - M_{1,1} = 0 \\ M_{0,1} + M_{1,0} = 0 \end{cases} \implies \begin{pmatrix} M_{0,0} \\ M_{1,1} \\ M_{0,1} \\ M_{1,0} \end{pmatrix} = u \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + v \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \quad (3.171)$$

On a deux opérateurs sur f et g :

$$\langle \nabla f, \nabla g \rangle, \quad \det(\nabla f, \nabla g) \quad (3.172)$$

3.6 Variance locale

3.6.1 Variance sur des fenêtres de l'image

La recherche de filtres quadratiques, pour différencier les phases osseuses et cellulaires dans les coupes immunohistologiques, nous a mené vers des opérateurs se rapprochant de la variance. Cependant, les matrices des parties précédentes qui interviennent dans le calcul des filtres deviennent difficilement inversibles numériquement lorsque la taille de la fenêtre augmente.

Nous sommes donc revenu à un calcul de variance statistique des couleurs de l'image sur des fenêtres de taille 20×20 pixels quadrillant notre image.

C'est à dire que sur chaque fenêtre, nous avons 400 valeurs rouges $(r_i)_{i \in [1, 400]}$,

$$V_{\text{fenêtre_rouge}} = \text{Var}((r_i)_i) = \frac{1}{400} \sum_i (r_i - \bar{r})^2 \quad (3.173)$$

et la variance totale est calculée en sommant les variances sur chaque couleur :

$$V_{\text{fenêtre}} = V_{\text{fenêtre_rouge}} + V_{\text{fenêtre_vert}} + V_{\text{fenêtre_bleu}} \quad (3.174)$$

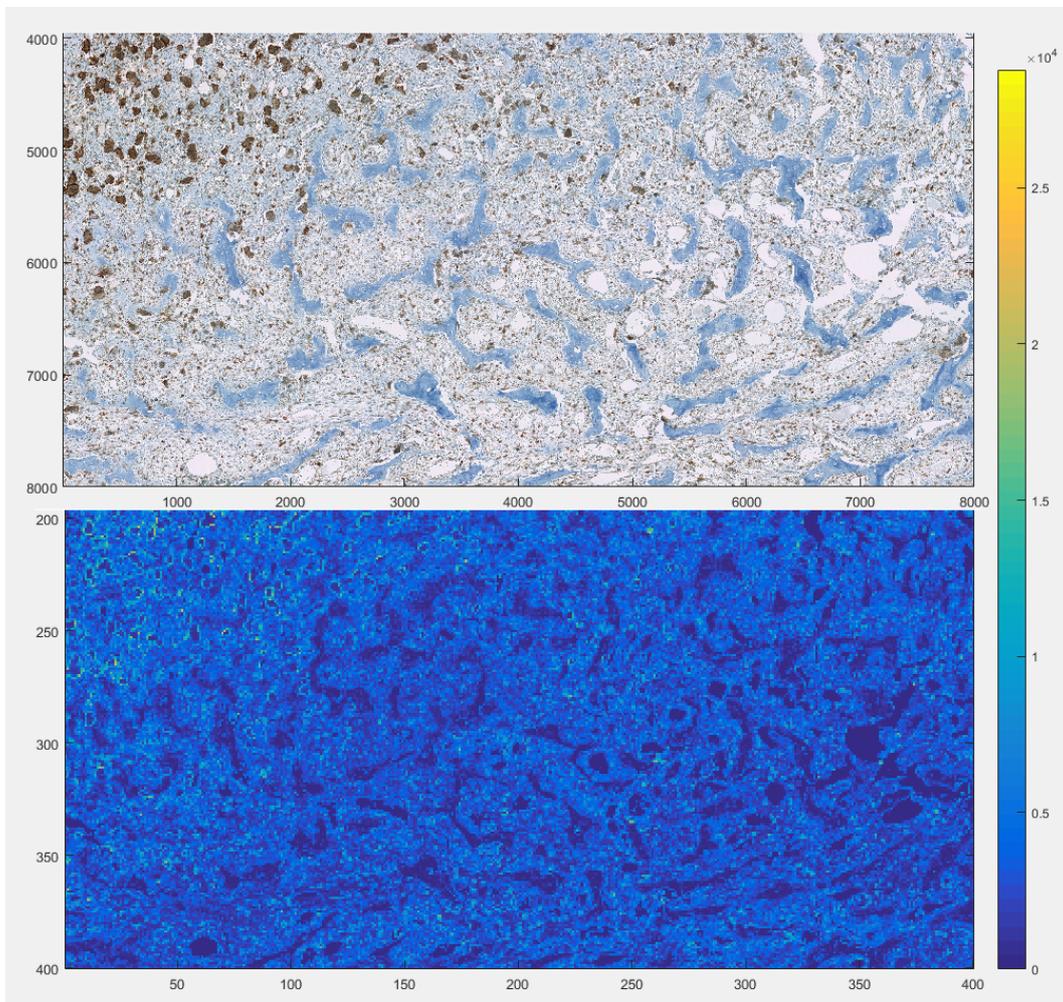


Figure 3.15: Image initiale au dessus, et variance des couleurs sur des fenêtres de tailles $20px \times 20px$ en dessous.

On observe une différence de variance sur les parties occupées par des cellules et celles recouvertes par de l'os. Comme pour tout le reste de la partie sur les filtres, le but n'est pas d'avoir une segmentation qui distingue tous les éléments biologiques de l'image, mais bien de se concentrer sur ce que le KNN n'arrive pas à classer correctement.

3.6.2 Variance locale lisse sur l'image

On peut définir un analogue local de la variance statistique.

Voici une expression de la variance :

$$\text{Var}(X) = \mathbb{E} \left((X - \mathbb{E}(X))^2 \right) \quad (3.175)$$

On peut décider de calculer une moyenne en pondérant les valeurs X_i avec un coefficient α_i où i parcourt l'image. La moyenne devient :

$$\mathbb{E}(X) = \frac{\sum_i \alpha_i X_i}{\sum_i \alpha_i} \quad (3.176)$$

Cette pondération peut nous permettre de donner plus de poids aux valeurs des couleurs proches d'un pixel, par exemple avec une gaussienne :

$$\mathbb{E}(X) = \frac{\sum_{i,j} \alpha_{i,j} X_{i,j}}{\sum_{i,j} \alpha_{i,j}} \quad (3.177)$$

$$\alpha_{i,j} = \exp \left(-\frac{1}{2\sigma^2} ((i - i_0)^2 + (j - j_0)^2) \right) \quad (3.178)$$

L'espérance devient ici un filtre gaussien, on peut donc utiliser le flou gaussien qui est disponible dans la plupart des langages de programmation et qui est optimisé. On a donc une variance locale sur une image qui serait :

$$\text{Var}(I) = \text{flou} \left((I - \text{flou}(I, \sigma))^2, \sigma \right) \quad (3.179)$$

Cette variance locale est un filtre quadratique.

Nous pouvons donc observer quelles sont les valeurs de cette variance locale sur une image d'immunohistologie dont nous avons sélectionné des parties osseuses et des cellules (figure 3.16).

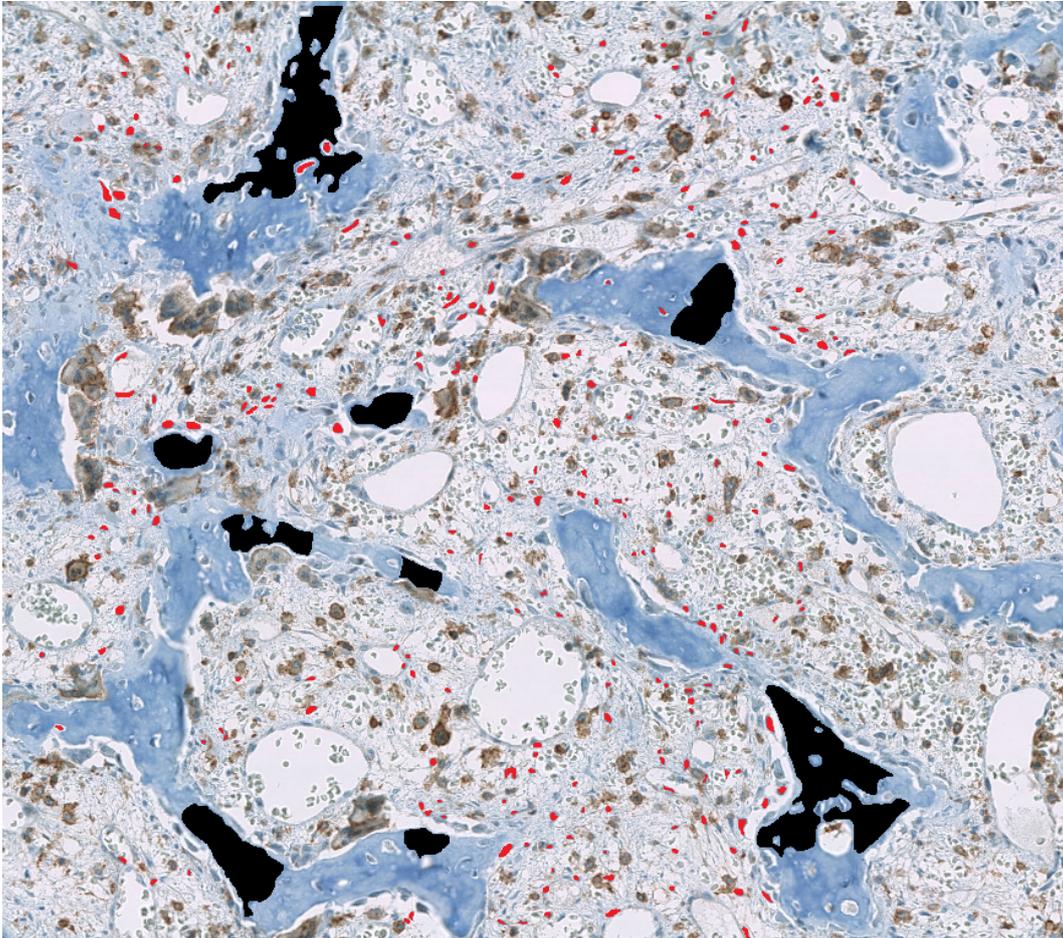


Figure 3.16: Échantillon d'os et de cellule sur une coupe d'immunohistologique.

Afin d'avoir des résultats plus visibles pour l'illustration, nous avons affiché un écart-type local en prenant la racine carrée. Nous obtenons une répartition des écart-types pour une valeur de $\sigma = 6,5$ sur les pixels des os ou des cellules sélectionnés qui sépare de manière satisfaisante ces deux classes (figure 3.17).

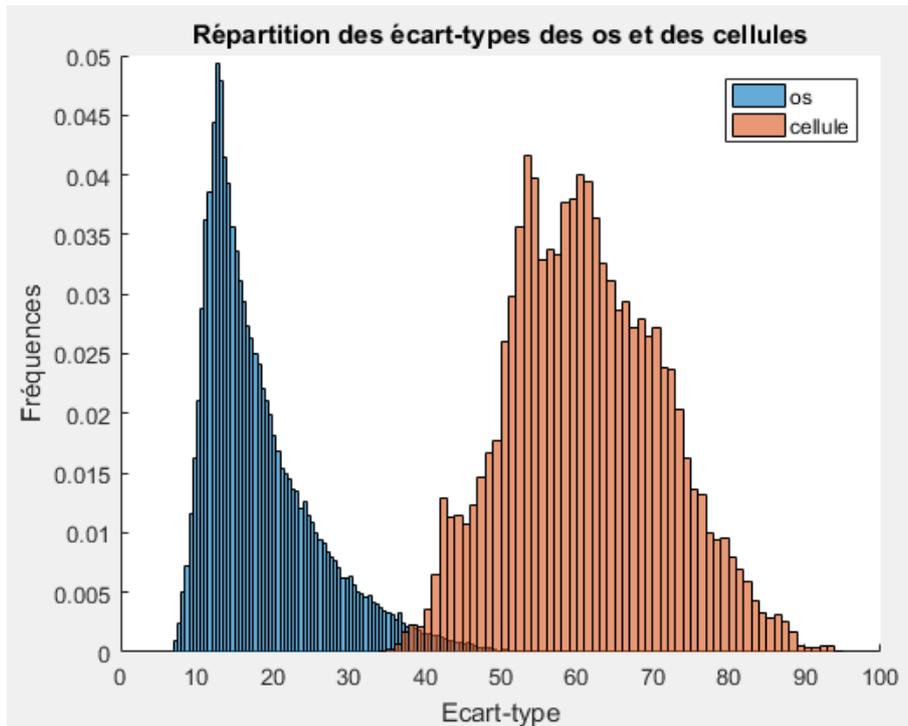


Figure 3.17: Répartition de l'écart-type entre l'os et les cellules.

Nous avons essayé de faire la même chose avec des moments d'ordre supérieur :

$$\mu_p = \mathbb{E}((X - \mathbb{E}(X))^p) \quad (3.180)$$

mais ils n'ont pas donné plus d'informations, même combinés avec la variance. Cependant, il semblerait que le fait de prendre des σ_1 et σ_2 distincts dans l'expression :

$$f(I) = \text{flou} \left((I - \text{flou}(I, \sigma_1))^2, \sigma_2 \right) \quad (3.181)$$

permettrait d'avoir de meilleurs résultats.

Le filtre non-linéaire ϕ peut s'écrire:

$$\phi(I) = f \left(\frac{1}{3} \sum_{i=1}^3 |I_i - f(I_i)|^p \right), \quad (3.182)$$

où I est l'image segmentée par KNN, et I_i avec $i = 1 \dots 3$, la matrice 2D des composantes du cube RGB, f le flou gaussien optimisé avec un rayon de $R = 6.5px$ et p un exposant obtenu par optimisation du score ou indice de Jaccard.

3.7 Conclusion

Dans ce chapitre, différentes façons de déterminer des filtres permettant la segmentation fine des coupes histologiques et immuno-histologiques a été proposée. Il n'y a pas de filtre idéal sur toutes les images et chaque filtre doit être utilisé pertinemment au cas par cas, par exemple en fonction de la nature dentelée ou en forme de travées de la phase osseuse. Le tableau 3.2 reprend la conclusion du chapitre 2 en ajoutant les scores associés à l'adjonction de filtres SVM ou de classification par variance locale. On voit qu'en moyenne le score est amélioré de 4 à 5% par ces approches. Une illustration des segmentations obtenues par ces différentes approches est

donnée figure 3.18. Dans l'exemple proposé, la variance locale donne les meilleurs résultats. En parallèle, ce chapitre détaille des analyses théoriques permettant la décomposition de tout filtre de voisinage $n \times m$, non interprétable directement, comme la SVM, sur des filtres aux dérivées partielles dans le cas d'images non bruitées et bruitées. En effet, pour envisager d'utiliser ces algorithmes dans un contexte d'application médicale, il est nécessaire de pouvoir interpréter les résultats obtenus. Une séparation par familles de problèmes sur une cohorte pourrait par exemple permettre l'utilisation de ces filtres aux dérivées partielles qui améliorent le temps de calcul relativement à la SVM et sont interprétables. De plus des familles de filtres linéaires, non-linéaires, homogènes et non homogènes, invariants par rotation et symétrie ont été déterminés. L'utilisation de ces filtres dans les itérations de la SVM non linéaire pourrait être envisagé pour appliquer des approches de haute performance en terme de score mais avec un gain considérable en temps de calcul.

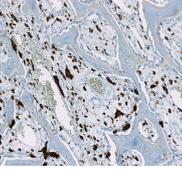
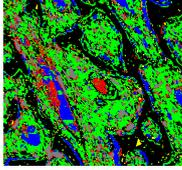
	KNN (supervised)
	
Parameters	k=15
CPU time	5.70 s
Score test	91.55 %
Score + lissage largeur 4	93.04 %
lissage 2 + variance locale ($p=0.2, s=11^p, r=6.5$)	96.21 %
lissage 2 + taille parties connexes ($s=30000$)	95.23 %
lisse + SVM($r=15$)	95.15 %

Table 3.2: Comparaison du score de Jaccard avec une approche KNN associée à différents filtres pour une coupe immunohistologique (marquage CD68). Les filtres explorés sont la variance locale, la discrimination sur la taille des parties connexes et l'utilisation d'un filtre SVM.

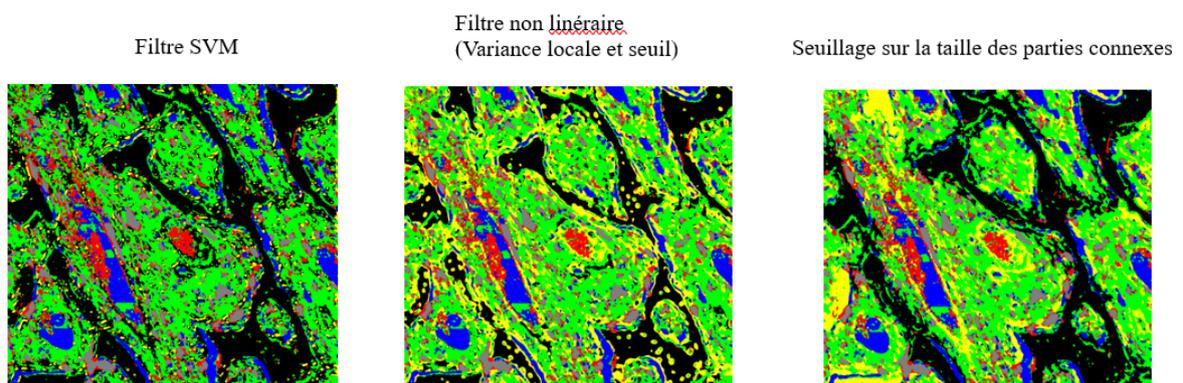


Figure 3.18: Applications des filtres SVM ($r = 15$), variance locale ($p = 0.2, s = 11^p, r = 6.5$) et retrait des parties connexes sur une coupe CD68 ($s = 30000$) et observations de leurs différences.

Chapter 4

Métriques et cartes de corrélation locale

Contents

4.1 Porosité et définition de longueurs caractéristiques	65
4.1.1 Définition et premiers calculs de porosité	65
4.1.2 Longueur caractéristique par méthode des fenêtres croissantes	67
4.1.3 Longueur caractéristique par méthode des semivariogrammes	68
4.1.4 Longueur caractéristique du pavage des cartes	70
4.2 Périmètre spécifique	71
4.3 Mesure locale des dimensions caractéristiques du tissu osseux . . .	73
4.4 Calcul de la densité cellulaire	75
4.5 Calcul du facteur de pertinence	78
4.6 Corrélations entre quantificateurs	78
4.6.1 Corrélations globales	79
4.6.2 Corrélations Locales	81
4.6.3 P-value et p-value locale	86

Dans ce chapitre, l'objectif est de proposer des méthodes de caractérisation géométriques des ostéosarcomes en proposant différentes méthodes de calculs de métriques adaptées aux images cliniques que nous étudions. À l'heure actuelle, il n'existe pas, à notre connaissance, de méthodes quantitatives permettant d'explorer les couplages entre réponse immunitaire, formation de tissu osseux ou fibreux et réponse au traitement. Pour explorer ce manque, les idées sous-jacentes sont dans un premier temps d'étudier la microarchitecture et, par la suite, de s'intéresser à la mécanique de ces milieux poreux complexes.

Dans ce chapitre nous proposons des méthodes pour la caractérisation de la porosité, du périmètre spécifique correspondant à la longueur de l'interface milieu osseux avec son environnement, de l'épaisseur des travées osseuses, de la densité cellulaire et du calcul des corrélation spatiales locales. Ceci constitue une des originalités de ce travail par rapport à la littérature du domaine. Les images explorées dans la suite sont segmentée par les méthodes décrites dans les chapitres 2 et 3.

4.1 Porosité et définition de longueurs caractéristiques

4.1.1 Définition et premiers calculs de porosité

La porosité ϵ est le rapport du volume de la phase non solide sur le volume total. Dans le contexte de cette étude, les images étant bidimensionnelles, il s'agira d'un rapport de surfaces.

$$\epsilon = \frac{V_{\text{pores}}}{V_{\text{total}}} \quad (4.1)$$

La porosité se calcule sur une image binarisée qui segmente la partie osseuse et le reste. Ici, on associe le label 1 à la “phase fluide” considérée comme la combinaison de tout ce qui n’est pas du tissu osseux. L’image binarisée est établie à partir de la segmentation, suivie d’un post-traitement : suppression de petites parties connexes de l’os et de petits trous dans l’os (ce post-traitement sera encore plus utile lors qu’il s’agira d’estimer les rayons moyens locaux des travées osseuses). Le calcul de la porosité sur l’image se fait en comptant les pixels ayant le label 1 et en divisant par le nombre total de pixels sur une région d’intérêt carrée.

La figure 4.1 montre une image de coupe histologique et sa binarisation à partir des méthodes de segmentation proposées dans le chapitre 2. La porosité globale sur cette image est de $\epsilon \approx 0,703$.

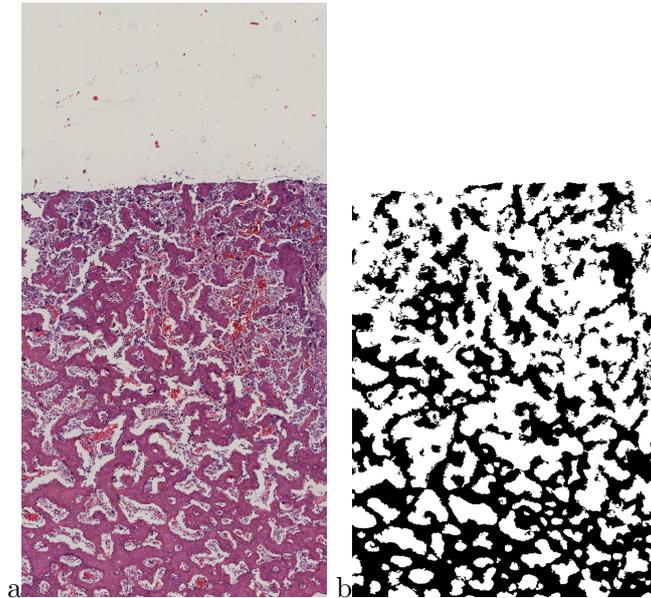


Figure 4.1: a) Coupe histologique H&E (taille 8928px*17568px, largeur pixel =0.466 μ m), b) image segmentée puis binarisée.

Cependant nous voyons que cette porosité globale a un sens limité sur cette image. Premièrement parce qu’il y a une partie de “vide” en haut et l’image, et deuxièmement parce que dans la partie où le tissu osseux est présent, la porosité peut varier localement. Nous allons nous intéresser à une estimation locale de la porosité. Une première méthode est de découper l’image en fenêtres carrées (ou pavage) et de calculer la porosité à l’intérieur de chaque fenêtre.

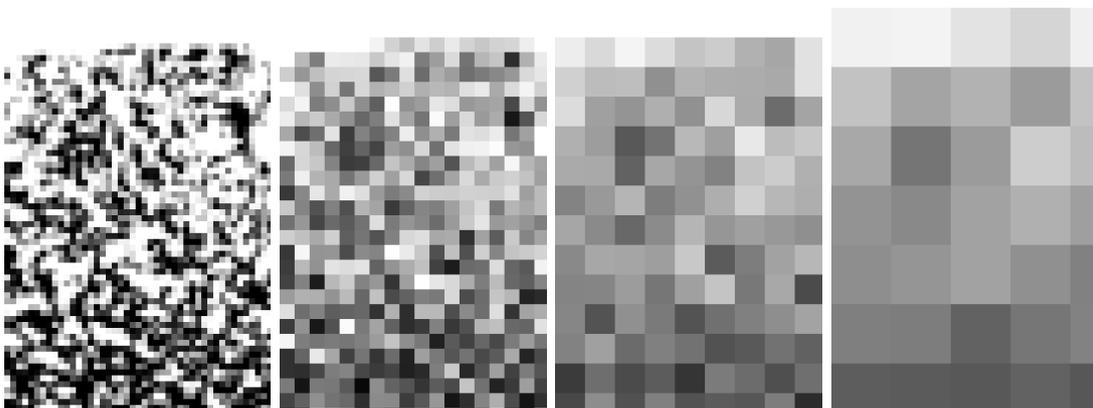


Figure 4.2: Illustration de la porosité sur des fenêtres avec un côté de 200, 500, 1000 puis 2000 pixels (blanc $\epsilon = 1$, noir $\epsilon = 0$).

Une largeur de fenêtre (ou région d’intérêt) trop petite ne rend pas bien compte d’une porosité

mais fait réapparaître les travées osseuses (figure 4.2) alors qu'une largeur trop importante peut faire perdre de l'information sur l'hétérogénéité spatiale du milieu. Il y a alors la nécessité de déterminer une taille caractéristique de la fenêtre avec laquelle nous allons estimer la porosité.

4.1.2 Longueur caractéristique par méthode des fenêtres croissantes

On peut observer l'évolution de la porosité sur des fenêtres de tailles croissantes et identifier à quelle taille de fenêtre le calcul de porosité est le plus pertinent. La figure 4.3 est un exemple sur l'image segmentée présentée (figure 4.1), avec une fenêtre dont le centre est au centre de l'image avec une largeur allant de 10 à 2000 par pas de 10. Nous observons une largeur à partir de laquelle la porosité atteint un plateau entre 700 et 800 pixels. L'existence de ce plateau traduit le fait que la porosité est quasi-stationnaire et il est alors possible de définir la porosité sur une fenêtre de taille caractéristique appartenant à ce plateau pour que cette région soit représentative du milieu. Ceci est vrai si cette propriété est vérifiée quelque soit le centre de la fenêtre. On parle alors de séparation d'échelles entre l'échelle du pore et l'échelle du tissu et une surface élémentaire représentative du milieu peut être définie [71].

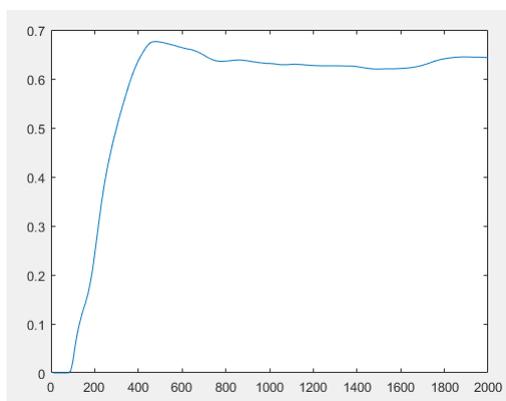


Figure 4.3: Calcul de la porosité sur des fenêtres carrées de tailles croissantes. En abscisse est représentée la largeur des fenêtres carrées et en ordonnée la porosité sur cette fenêtre.

Pour détailler l'évolution de la courbe 4.3, on observe que pour une fenêtre trop petite (figure 4.4), on est encore dans une partie segmentée avec le label 0 (partie osseuse), c'est pourquoi la porosité est de 0 pour un petit "rayon" (ou largeur de fenêtre). Ensuite, la phase "fluide" est prise en compte pour recouvrir une partie plus représentative de l'image.

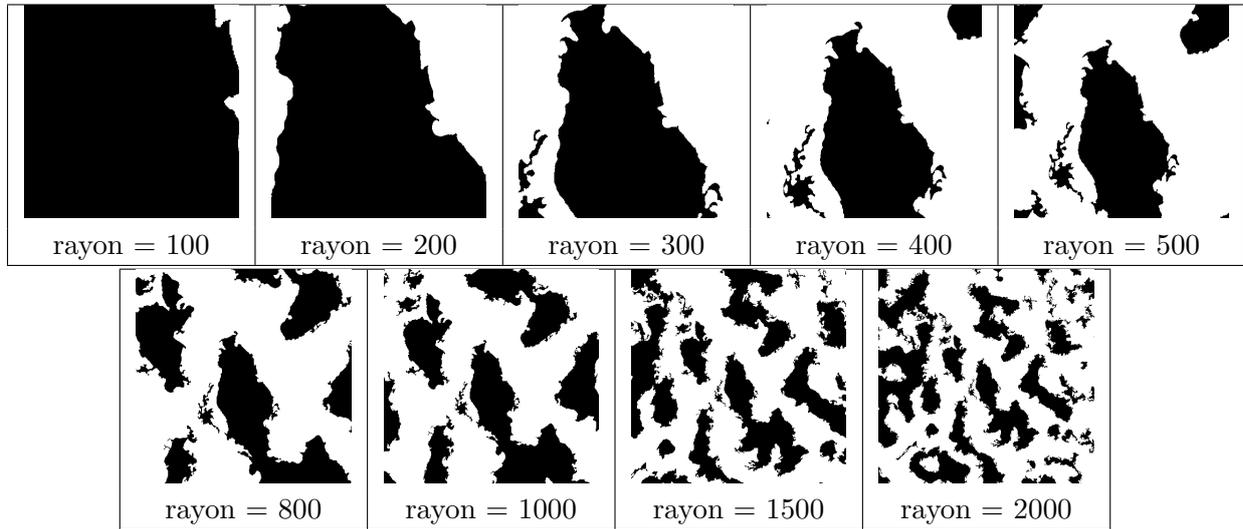


Figure 4.4: Images binarisées à partir de segmentation KNN. La phase osseuse est identifiée en noir, le reste en blanc. Zooms sur l'image binarisée en fonction du "rayon" ou largeur de fenêtre, ramenés à la même surface.

Dans le cas le cas de l'ostéosarcome, la phase osseuse est cependant très hétérogène spatialement et en fonction de la position du centre de la fenêtre il est parfois difficile d'identifier un plateau comme l'illustre la figure 4.5. Il n'y a donc pas séparation d'échelles dans ce milieu.

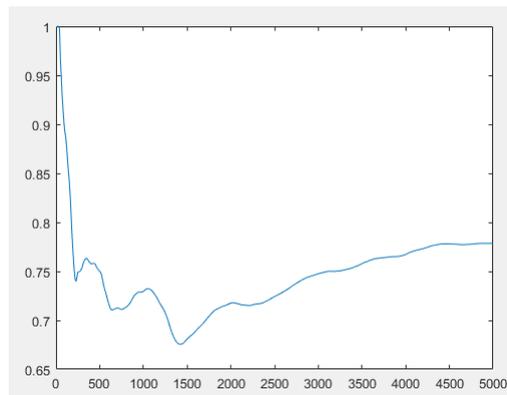


Figure 4.5: Distribution de la porosité pour des fenêtres de tailles croissantes centrées en pixel (6784, 4464).

4.1.3 Longueur caractéristique par méthode des semivariogrammes

Une autre méthode pour déterminer une distance caractéristique est l'utilisation de semi-variogrammes [72].

On définit la semi-variance γ où X représente la valeur binaire 0 ou 1 du pixel dans la position x , h est une distance:

$$\gamma(h) = \frac{1}{2} \mathbb{E}(|X(x+h) - X(x)|^2) \quad (4.2)$$

c'est un écart quadratique entre le profil binaire et ce même profil translaté.

Nous avons alors :

$$\gamma(0) = 0 \quad (4.3)$$

Supposons que h soit de norme suffisamment grande, $\gamma(h)$ va se rapprocher d'une valeur telle que $X(x+h)$ et $X(x)$ soient indépendants. On a alors :

$$\mathbb{E}(|X(x+h) - X(x)|^2) = \mathbb{E}(|X(x+h) - X(x)|) = \mathbb{P}(X(x+h) \neq X(x)) \quad (4.4)$$

$$= \mathbb{P}(X(x+h) = 1)\mathbb{P}(X(x) = 0|X(x+h) = 1) + \mathbb{P}(X(x+h) = 0)\mathbb{P}(X(x) = 1|X(x+h) = 0) \quad (4.5)$$

$$= \mathbb{P}(X(x+h) = 1)\mathbb{P}(X(x) = 0) + \mathbb{P}(X(x+h) = 0)\mathbb{P}(X(x) = 1) \quad (4.6)$$

$$= 2\varepsilon(1 - \varepsilon) \quad (4.7)$$

où ε est la porosité. Donc, pour h de norme grande

$$\gamma(h) \xrightarrow{\|h\| \rightarrow +\infty} \varepsilon(1 - \varepsilon) \quad (4.8)$$

Le profil des semi-variances dans une direction donnée est un semi-variogramme. Voici un exemple de semi-variogramme sur une image binaire segmentant l'os et la phase fluide dont chaque courbe correspond aux directions $(1, 0)$, $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, $(0, 1)$ et $(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$:

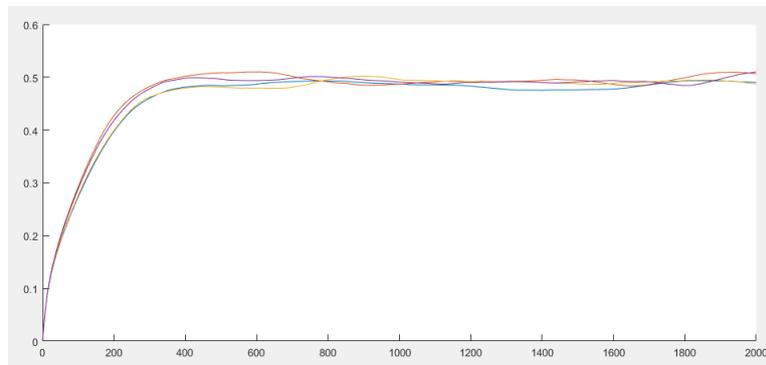


Figure 4.6: Semivariogramme indiquant γ en fonction de h de l'image binaire, les couleurs des quatre courbes sont associées à des directions de h différentes. 4.1b.

On observe sur le semivariogramme 4.6 qu'il ne semble pas y avoir d'anisotropie, et qu'à partir d'une distance de 800 pixels, on peut considérer que $X(x)$ et $X(x+h)$ sont indépendants. Ce qui est consistant avec la méthode des fenêtres croissantes. Cependant, cette approche populaire en géophysique et méthodes de krigeages pour estimer les caractéristiques d'un milieu poreux défini par des informations parcellaires, a tendance à filtrer les variations de la moyenne des variables et dans le but d'étudier l'hétérogénéité de l'ostéosarcome, la méthode des fenêtres croissantes sera préférée.

Le demivariogramme 4.7 montre un exemple de surface de γ en fonction de h sur un domaine de \mathbb{R}^2 . L'observation des lignes de niveaux de γ permet de visualiser l'invariance ou la dépendance par rotation de γ .

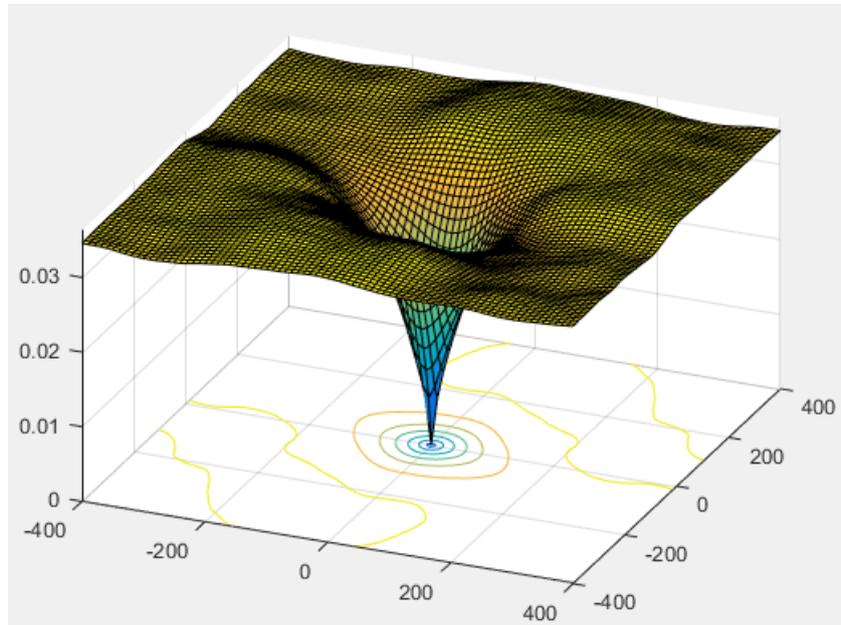


Figure 4.7: Semivariogramme sur h parcourant $[-400, 400] \times [-400, 400]$ sur l'image binarisée associée à l'os une coupe immunohistologique marquée au CD8. Avec représentation des lignes de niveau en bas.

4.1.4 Longueur caractéristique du pavage des cartes

Bien qu'aucune longueur caractéristique ne puisse être définie sur l'ensemble de la lame, nous nous inspirons de l'approche numérique de Durlfsky [47] pour calculer des propriétés de transport dans des milieux poreux, par méthode de pavage, où une séparation d'échelle ne peut pas être définie. Dans ce contexte un pavage où les éléments peuvent prendre une taille arbitraire est réalisé. Dans le contexte de cette étude une taille caractéristique de 800 pixels est choisie. La carte de porosité sur coupe entière immunohistologique est illustrée figure 4.8. Les zones principalement cellulaires sont représentées en vert et les zones principalement osseuses en noir.

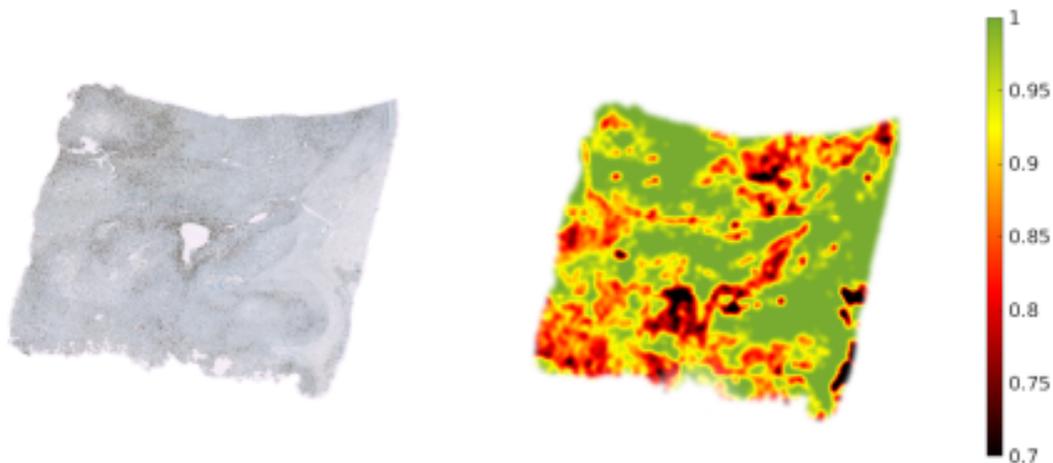


Figure 4.8: a) Coupe immunohistologique entière marquée au CD68 (largueur de coupe 4cm avec une résolution de $0.466 \mu m$), b) carte de porosité sur un pavage dont les éléments ont une taille caractéristique de $800px$.

4.2 Périmètre spécifique

Afin de caractériser le tissu osseux géométriquement, on souhaite estimer la longueur de l'interface ou périmètre entre la phase fluide et la phase osseuse. En effet, en plus de donner une mesure de l'interface entre la phase osseuse et la phase liquide, le périmètre comparé à la porosité donne des informations sur la régularité de cette interface. On définit qu'un pixel appartient à l'interface, si un pixel associé à une label 0 ou 1 possède au moins un pixel voisin associé à l'autre label. La frontière entre ces deux interfaces n'est pas rigoureusement définie, mais on souhaite se rapprocher du périmètre d'une courbe "théorique" qui séparerait les deux phases. On suppose que cette courbe possède une longueur au sens de mesure de Lebesgue, mais nous verrons plus tard quelle possède un aspect fractal et que le "périmètre" peut dépendre de l'échelle à laquelle on s'intéresse.

La première approche "grossière" consiste à effectuer un comptage de pixels qui cependant surestime la longueur théorique comme illustré figure 4.9. Nous voyons qu'entre l'illustration centrale et celle de droite, il y a un facteur $\sqrt{2}$ sur l'estimation du même périmètre par comptage de pixels, qui se maintiendra quel que soit l'échelle considérée. Il y a donc un biais en fonction de l'orientation alors que le périmètre est une quantité invariante par rotation.

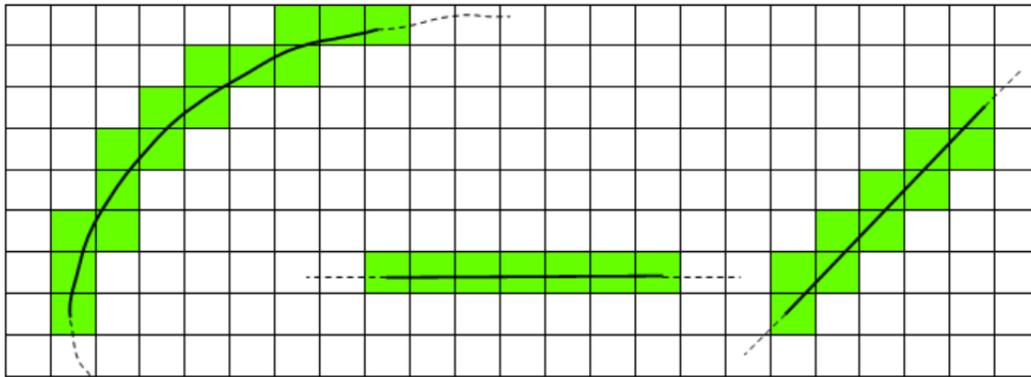


Figure 4.9: Illustration de l'erreur commise sur le calcul du périmètre par comptage de pixels.

Afin de réduire ce biais, il est possible de dilater la frontière avec des disques de rayon r . La figure 4.10 montre un exemple de dilatation d'une frontière, sur une partie de l'image figure 4.1b et c, avec des disques de rayon $r = 5$ et $r = 20$.

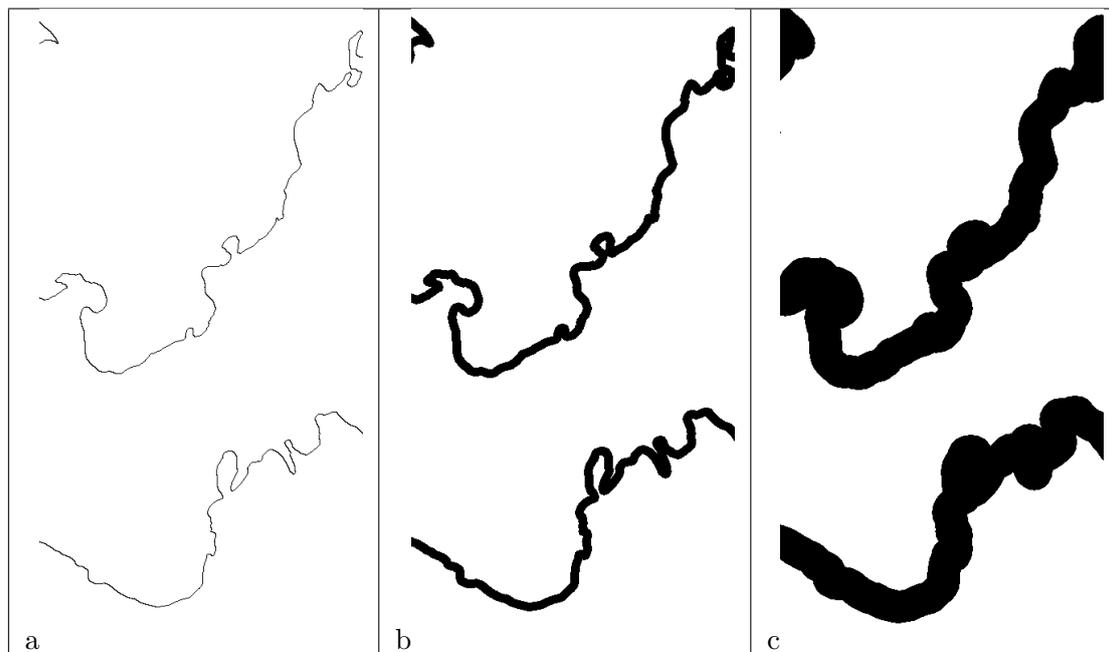


Figure 4.10: a) Frontière entre les labels, b) dilatation avec un rayon de 5 pixels et c) dilatation avec un rayon de 20 pixels.

On compte le nombre de pixels de cette dilatation qui vaut approximativement $2r \times p$, où p est le périmètre. On fait ensuite tendre r vers 0 et on estime la limite du quotient du nombre de pixels de cette dilatation et $2r$ qui est p . Théoriquement, la dilatation F_r de la frontière est donnée par :

$$F_r = \{x \in \Omega, \text{dist}(x, F) \leq r\} \quad (4.9)$$

où x est la coordonnée d'un élément de la frontière Ω .

Le contenu de Minkowski (Formule de Steiner-Minkowski) dit que pour une courbe paramétrée compacte dans \mathbb{R}^2 de classe C^2 , alors sa longueur est la limite :

$$\frac{1}{2r} \mu(F_r) \xrightarrow{r \rightarrow 0} p \quad (4.10)$$

où μ est ma mesure de Lebesgue.

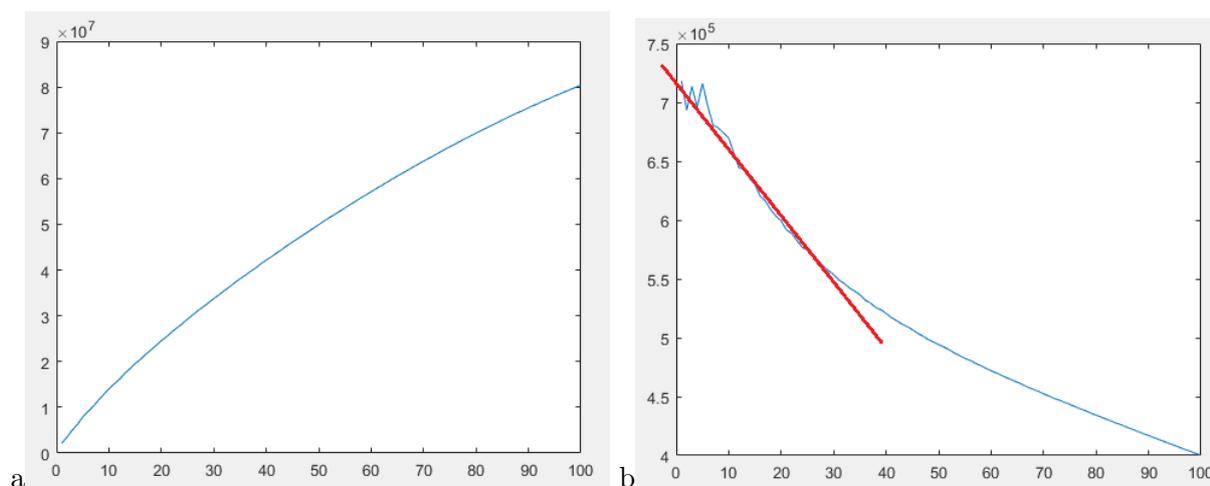


Figure 4.11: a) Nombre de pixels de la dilatation de la frontière en fonction du rayon r des disques, b) régression linéaire du quotient du nombre de pixels de la dilatation de la frontière et $2r$, en fonction de r .

La figure 4.11a illustre le nombre de pixels de la dilatation de la frontière en fonction du rayon r des disques, la figure 4.11b illustre le quotient du nombre de pixels et $2r$ en fonction de

r . Sur cette dernière courbe bruitée proche de 0, on peut effectuer une régression linéaire.

Sur la figure 4.11b, nous pouvons ainsi estimer un périmètre spécifique d'environ $7,2 \times 10^5$ pixels entre la phase fluide et osseuse.

4.3 Mesure locale des dimensions caractéristiques du tissu osseux

Dans cette partie, nous allons chercher à déterminer des longueurs caractéristiques du tissu osseux et en particulier l'épaisseur des travées. Dans cette section l'image considérée est une coupe H&E, segmentée par KNN avec retrait des petites parties connexes (voir chapitre 2).

A l'intérieur de l'os, la distance de chaque pixel à l'interface est calculée (figure 4.12) en prenant la distance euclidienne relative au pixel de l'interface le plus proche.

Cette méthode d'approximation des largeurs des travées osseuses se rapproche de l'utilisation des lignes de partage des eaux qui a déjà montré son intérêt dans l'étude morphologique de l'os [73].

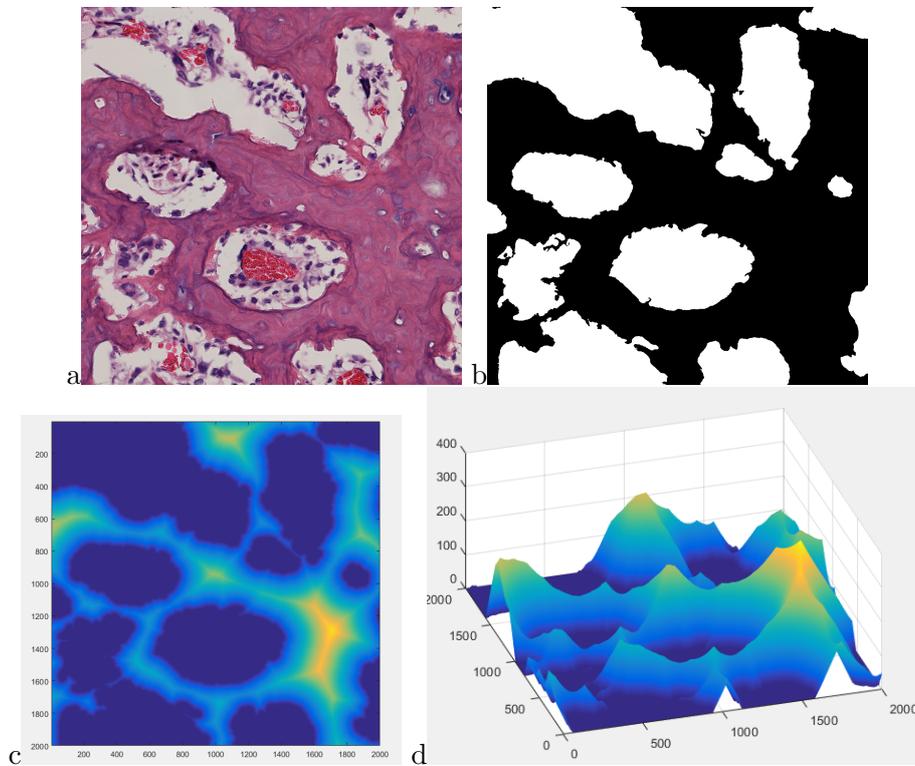


Figure 4.12: a) Zoom sur coupe histologique marquée H&E, b) Image binaire associée avec retrait des petites parties connexes, c) distance de chaque pixel de la phase osseuse à l'interface os/fluide en représentation planeaire (légende équivalente à l'axe vertical de la figure d), d) distance de chaque pixel de la phase osseuse à l'interface os/fluide en représentation.

On peut interpréter cette distance, d'un pixel de l'os à l'interface, comme le plus grand rayon des disques contenus dans la partie osseuse (figure 4.13a). Ainsi, les maxima locaux de la distance à l'interface os/fluide sont les centres des cercles tels que localement, il est impossible de déterminer de plus grand cercle. La figure 4.13b montre les maxima locaux de la distance obtenus sur l'image binaire.

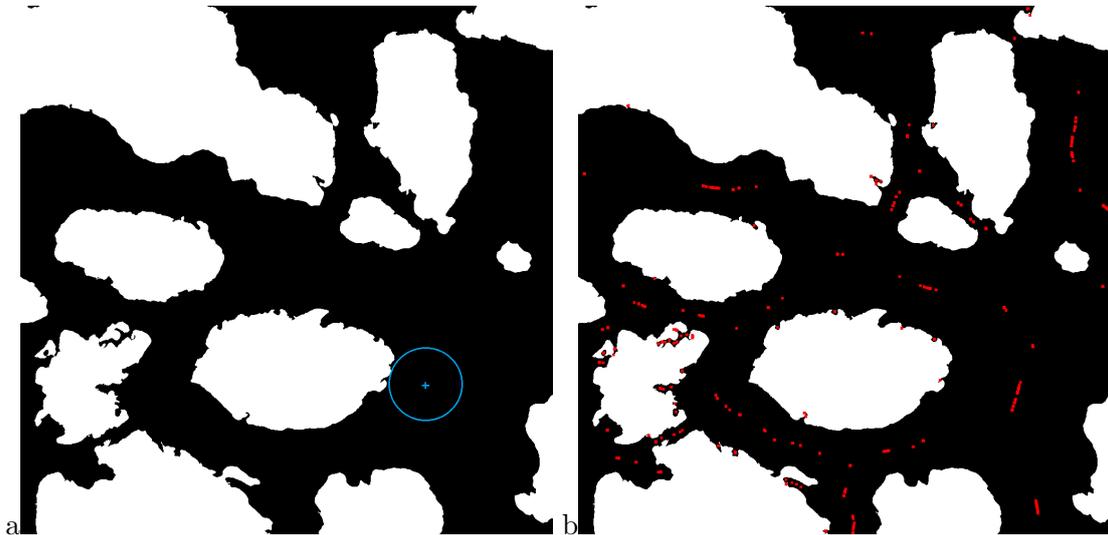


Figure 4.13: a) Plus grand cercle contenu dans l'os centré en un pixel fixé représenté par un point bleu, b) maxima locaux de la distance os/interface marqués par des points rouges.

Après avoir obtenu des maxima locaux sur lesquels identifier les distances, toutes ces valeurs de maxima ne sont pas à considérer avec le même poids pour définir une taille ou rayon caractéristique du tissu osseux. Nous allons utiliser le watershed dans la suite afin d'associer une partition de l'os à l'ensemble des pixels sur lesquels un maximum local est identifié. Le watershed peut-être vu comme une descente de gradient.

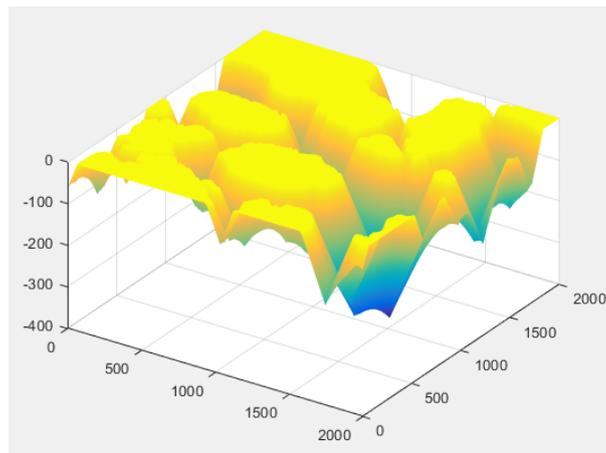


Figure 4.14: Opposé de la distance os/interface.

La figure 4.14 représente la surface qui, à chaque pixel, donne le négatif de la distance à la partie segmentée comme fluide. Dans cette figure, un pixel de l'os est associé à un minimum local, correspondant aux maxima locaux donnés dans la figure 4.13b. La partie osseuse est labellisée en associant un même label à chaque pixel associé au même minimum local par la méthode watershed, figure 4.15. L'aire de la zone associée à un minimum local (représentation par chaque couleur) donne un poids avec lequel considérer le maximum local de la distance dans le tissu osseux (figure 4.16).

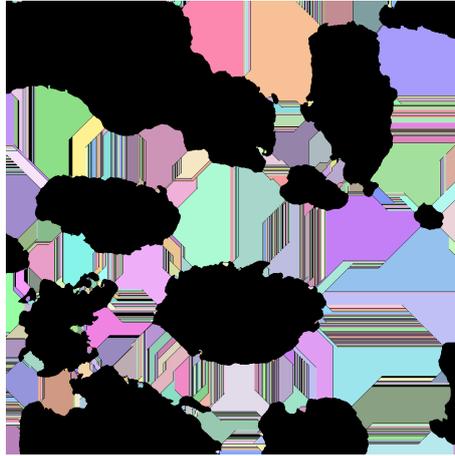


Figure 4.15: Partition de l'os obtenue par l'algorithme du watershed.

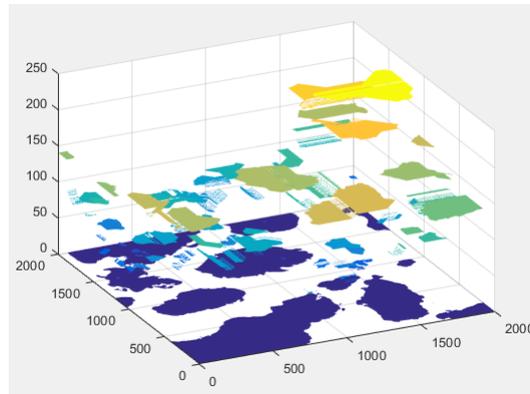


Figure 4.16: Distance maximale (axe vertical) sur chaque partie obtenue par le watershed. Les axe horizontaux correspondent aux variables spatiales.

Nous pouvons ainsi déterminer une épaisseur moyenne des travées osseuse. Sur l'image explorée dans ce paragraphe, la moyenne des maxima locaux, pondérée par la surface de la zone associée, donne un rayon caractéristique moyen de:

$$r \approx 145,03 \text{ pixels} \quad (4.11)$$

Ce processus peut être défini sur un pavage de la coupe entière, de manière analogue au calcul de porosité, afin d'obtenir une cartographie des tailles caractéristiques des travées osseuses.

4.4 Calcul de la densité cellulaire

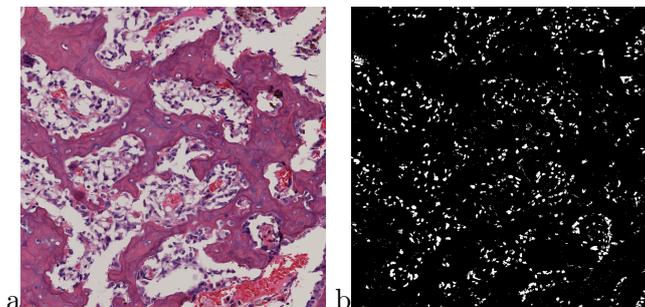


Figure 4.17: a) Coupe histologique H&E, b) image binaire issue de la segmentation KNN avec filtre SVM indiquant les cellules en blanc.

Dans la suite, on utilise la segmentation obtenue avec le KNN et l'adjonction d'un filtre SVM, pour avoir une image binaire qui marque les pixels classés dans les cellules avec le label 1, le

reste étant défini par le label 0 (figure 4.17) . La recherche de chaque cellule sur l'image binaire correspond à la recherche des parties connexes. On retire les parties connexes qui contiennent un nombre de pixels en dessous d'un certain seuil ($50px$), correspondant au bruit. La figure 4.18 montre la répartition des tailles (en pixels) des parties connexes obtenues. Dans cet exemple 5496 parties connexes sont obtenues.

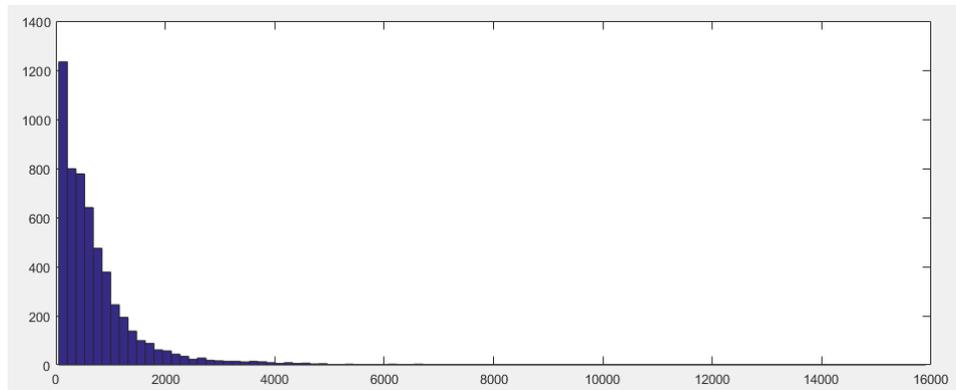


Figure 4.18: Répartition de la taille des parties connexes (cellules) de plus de 50 pixels. Ordonnées: nombre de parties connexes, abscisses: surface en pixels.

Par la suite, on peut déterminer le barycentre (x_G, y_G) des parties connexes qui se calcule avec la moyenne des positions des pixels de la partie connexe, c'est l'analogie discret du calcul de centre de gravité dans le cas continu:

$$x_G = \frac{1}{|C|} \int_C x d\tau, \quad y_G = \frac{1}{|C|} \int_C y d\tau \quad (4.12)$$

La figure 4.19a montre la répartition des barycentres des cellules dans l'image.

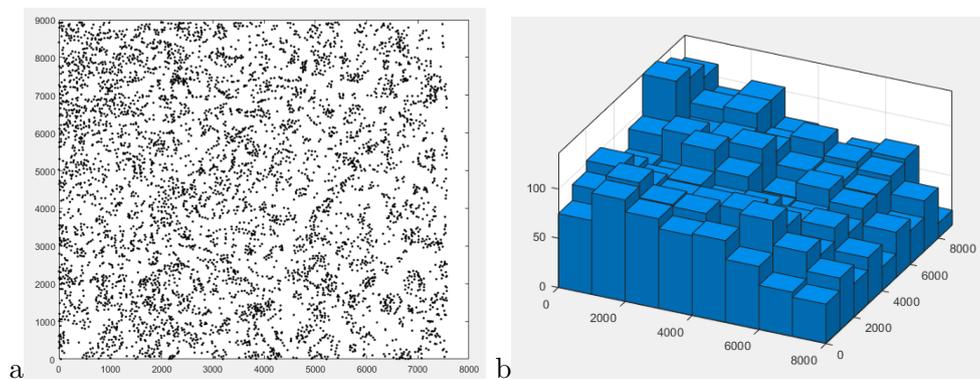


Figure 4.19: a) Répartition des barycentres des cellules dans l'image, b) histogramme de la répartition des centres des cellules.

On peut ainsi une à une identifier à quelle partie de l'image correspond la i -ième partie connexe et identifier le barycentre de chaque cellule. Par exemple, voici la 1690 ième marquée par un carré vert en son centre (figure 4.20).

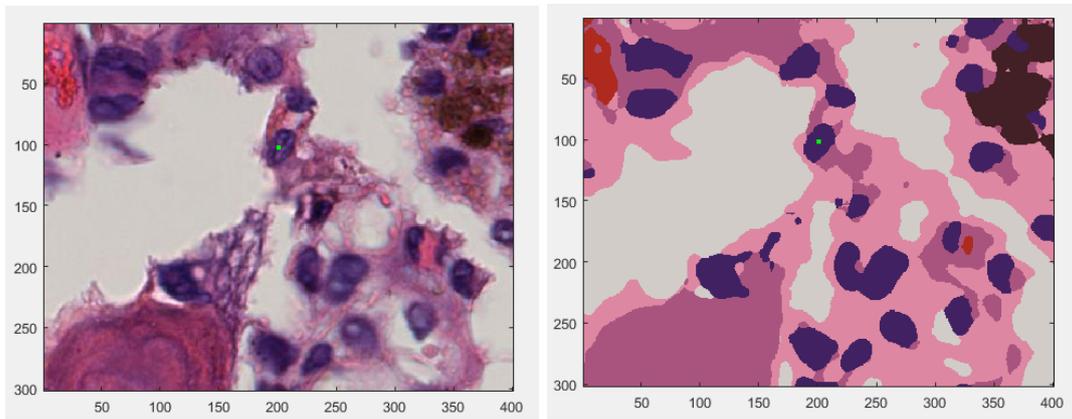


Figure 4.20: Centre de gravité d'une cellule (marqué par un carré vert) sur l'image de départ et l'image segmentée.

Dans la suite, deux points de vue seront pris en compte: on peut considérer uniquement les centres des cellules (c'est-à-dire uniquement leur présence), ou alors on peut pondérer ces centres avec l'aire. Les deux approches sont importantes et doivent être utilisées. Le fait de pondérer par l'aire de la cellule rend les statistiques sur celles-ci moins sensibles aux paramètres choisis pour la segmentation. La figure 4.21 montre la répartition des cellules mais avec l'aire représentée sur l'axe des z .

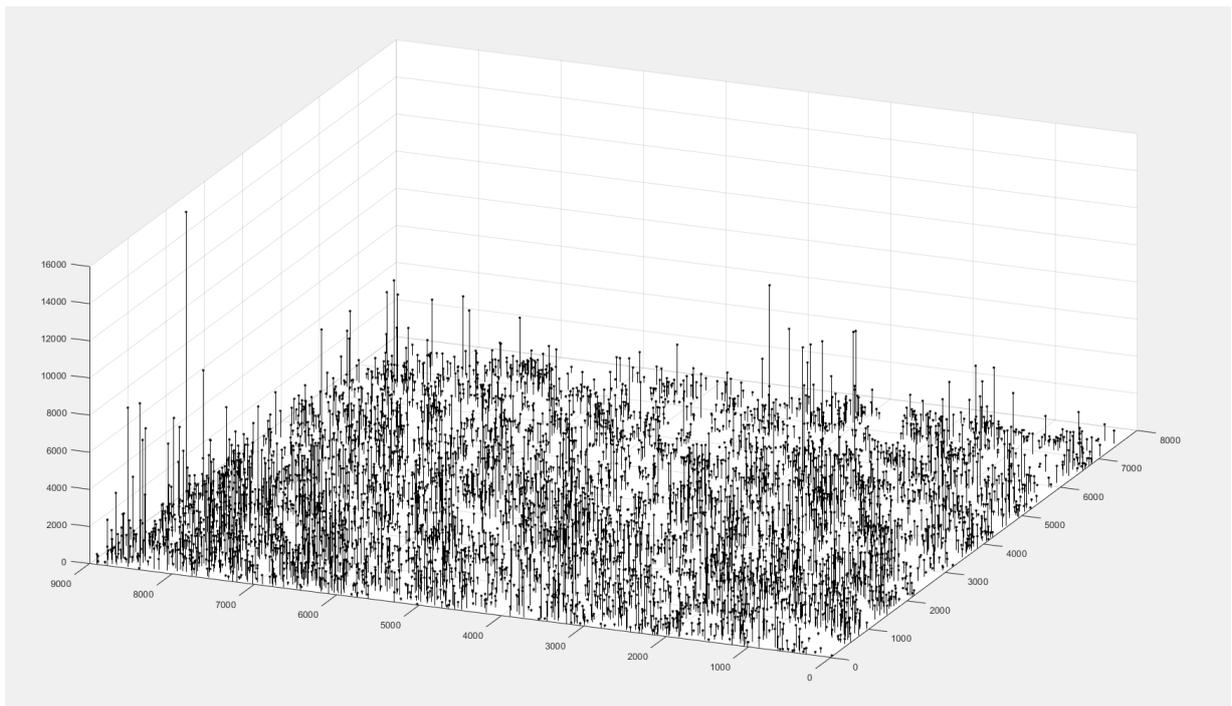


Figure 4.21: Répartition des centres des cellules dans l'image avec l'aire sur l'axe des z .

De la même manière que nous calculons la porosité du milieu, nous pouvons calculer la densité des cellules sur un pavage de l'image binarisée, figure 4.22. Cette densité cellulaire peut être basée sur la surface ou sur le nombre de cellules sur une région d'intérêt. Dans la suite les densités cellulaires explorées calculent le nombre de cellules par unité de surface. Pour s'affranchir des biais liés à la présence d'une phase osseuse dans l'ostéosarcome, seule la phase fluide est considérée comme surface d'exploration. Ainsi la densité cellulaire D_c donnée en mm^{-2} sur un élément du pavage, est définie par:

$$D_c = \frac{\text{nombre de cellules}}{\text{surface de la phase non osseuse}} \quad (4.13)$$

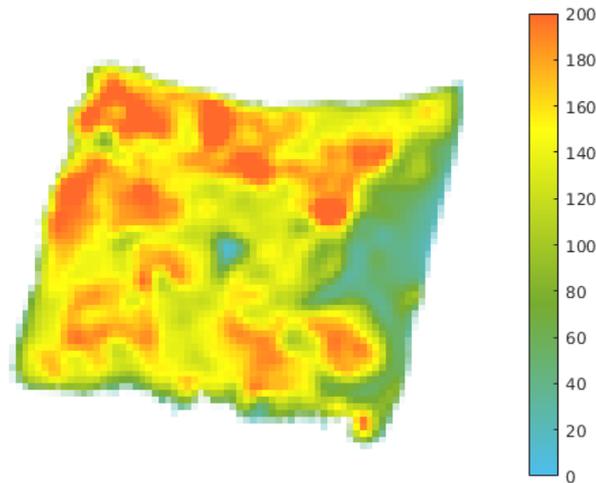


Figure 4.22: Carte de densité cellulaire avec pavage sans recouvrement de taille 800 pixels \times 800 pixels pour le marquage CD68, correspondant à la coupe 4.8a.

4.5 Calcul du facteur de pertinence

Afin d'éviter tout biais lors de la prise en compte de régions externes au tissu ou de régions de tissu déchirées, un facteur de pertinence a été implémenté pour tous les calculs de cartes et est illustré sur la figure 4.23. Pour l'évaluer, un masque est créé basé sur les grandes régions vides connectées. La pertinence est le pourcentage de l'élément de maillage qui est réellement occupé par le tissu. Dans la suite nous utiliserons le facteur de transparence dans l'illustration des données pour traduire ce facteur de pertinence, c'est à dire que lorsque le facteur de pertinence est 0, le fond sera blanc et au contraire lorsque le facteur de pertinence est 100, le masque sera complètement transparent.

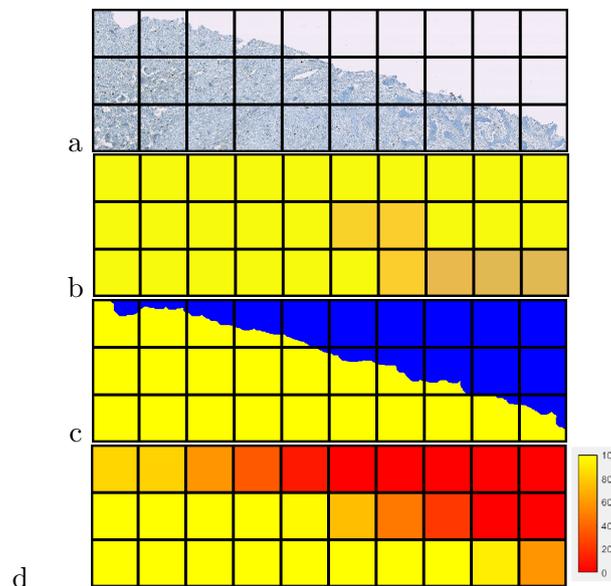


Figure 4.23: a) Pavage sur un zoom de l'image immunohistologique marquée CD8, b) porosité calculée sur chaque élément du pavage, c) masque du tissu en bleu, et d) facteur de pertinence en % multiplié par la porosité allant de 0 à 1).

4.6 Corrélations entre quantificateurs

Dans cette section paragraphe, nous proposons de développer une technique permettant d'identifier les corrélations entre porosité et densités de populations cellulaires d'intérêt. Cette technique

visé à obtenir ces coefficients de corrélation de manière spatialisée au lieu de calculer un coefficient global sur l'ensemble de la lame.

4.6.1 Corrélations globales

La première approche que nous avons mise en oeuvre pour chercher une corrélation entre porosité et densité cellulaire est le calcul de corrélation globale de type coefficient de corrélation de Pearson sur les données vectorisées de porosité et de densité cellulaire. En testant cette approche, sur un patient et pour trois marquages histologiques représentés figure 4.24, et en étudiant les corrélations entre soit la porosité soit la largeur des travées et différentes mesures de la densité cellulaire (densité des cellules marquées, densité des cellules totales, pourcentage de cellules marquées par rapport aux cellules totales), les coefficients de corrélations 4.25 restent de faible amplitude (au maximum 0.256), ce qui rend difficile les conclusions possibles quant aux corrélations entre la microarchitecture osseuse et la densité des populations cellulaires. Dans les images de la figure 4.24 nous pouvons remarquer que les images ont été recalées spatialement. Dans le cadre de cette thèse différentes méthodes de recalage ont été explorées pour étudier par exemple les corrélations entre différentes populations cellulaires, et ce travail est développé en annexe A pour préserver la cohérence de ce chapitre.

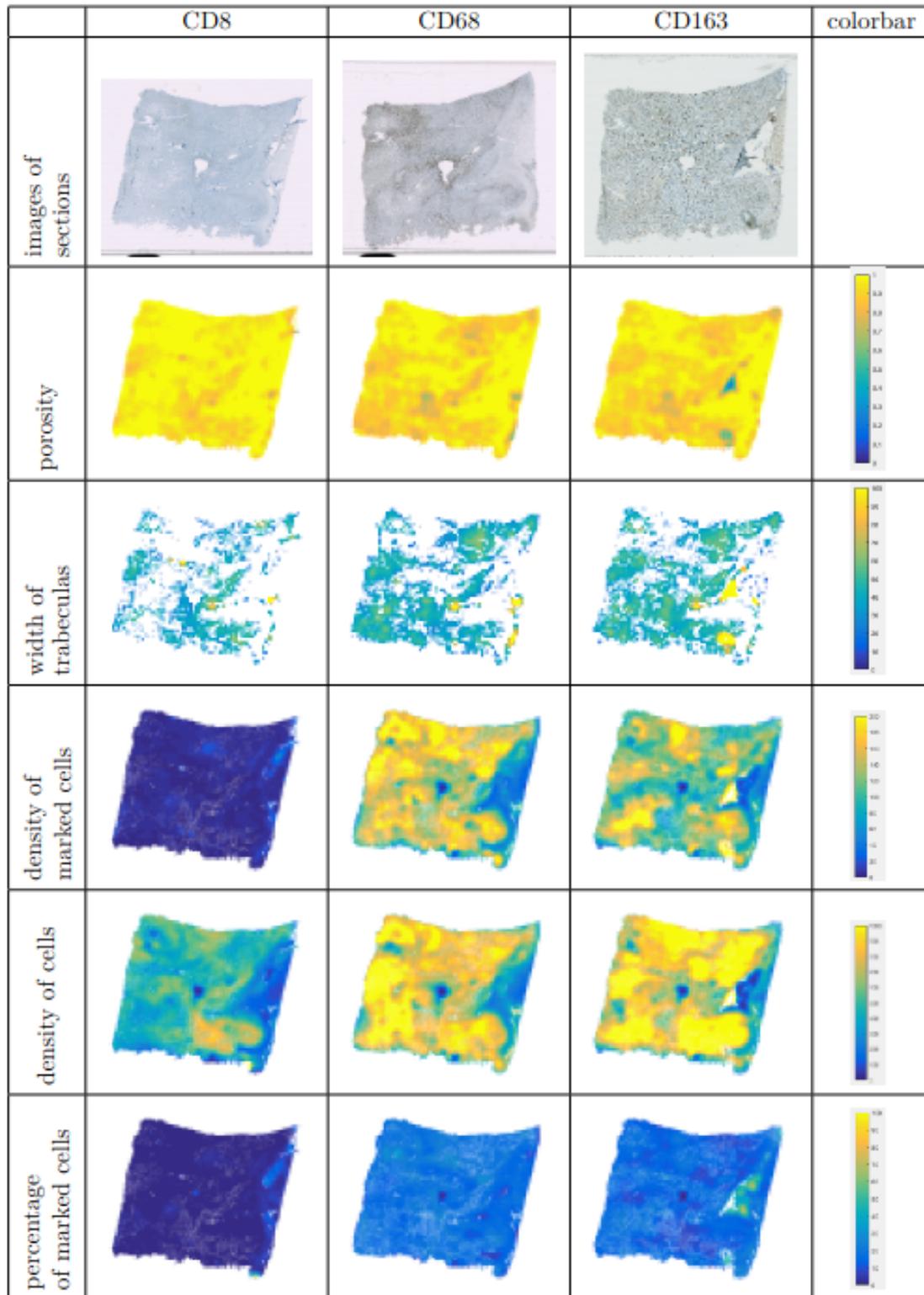


Figure 4.24: Chaque ligne de tableau représente dans l'ordre la coupe entière immunohistologique originale, la porosité, la largeur des travées, la densité en cellules marquées, la densité de toutes les cellules et le pourcentage de cellules marquées relativement aux cellules totales pour les marquages CD8, CD68 et CD163.

	S_{CD8}	C_{CD8}	Pe_{CD8}	S_{CD68}	C_{CD68}	Pe_{CD68}	S_{CD163}	C_{CD163}	Pe_{CD163}
P_{CD8}	-0.072	-0.134	-0.033	-0.06	-0.115	0.042	-0.041	-0.033	0.005
W_{CD8}	0.256	-0.111	0.142	-0.102	-0.081	-0.071	0.057	-0.099	0.131
P_{CD68}	0.023	-0.03	0.04	-0.116	-0.206	0.059	-0.064	-0.043	0.065
W_{CD68}	0.083	-0.247	0.117	-0.117	-0.077	-0.111	-0.066	-0.196	0.122
P_{CD163}	0.072	0.029	0.049	0	-0.064	0.064	-0.207	-0.065	-0.031
W_{CD163}	-0.02	-0.24	0.018	-0.209	-0.213	-0.066	0.209	-0.012	0.136

Figure 4.25: Coefficients de corrélation de Pearson globaux, P : porosité, W : largeur des travées, S: densité des cellules marquées, C: densité des cellules totales, Pe : pourcentage de cellules marquées relativement aux cellules totales

4.6.2 Corrélations Locales

Pour pallier aux limites d'un calcul de coefficient de corrélation global quant à son exploitation possible du point de vue clinique, nous avons cherché à développer une approche locale. Pour définir un coefficient de corrélation local, il faut tout d'abord définir une moyenne locale. Il y a une contrainte de distance caractéristique lorsqu'on rend locale une quantité, nommons cette distance l . Une quantité peut être rendue locale par calcul sur des fenêtres dont la largeur est la distance caractéristique L , mais également sur des fenêtres plus petites associées à des zones glissantes de taille caractéristique l .

Par exemple, dans la figure 4.26, on peut voir l'estimation locales de la moyenne d'un signal. En bleu, des fenêtres de taille L , en rouge des fenêtres plus petites associées à des zones de taille l et un schéma montrant la version 2D de ce calcul de moyennes à l'aide de fenêtres. On identifie sur cette figure que l donne une approximation plus précise de la moyenne du signal.

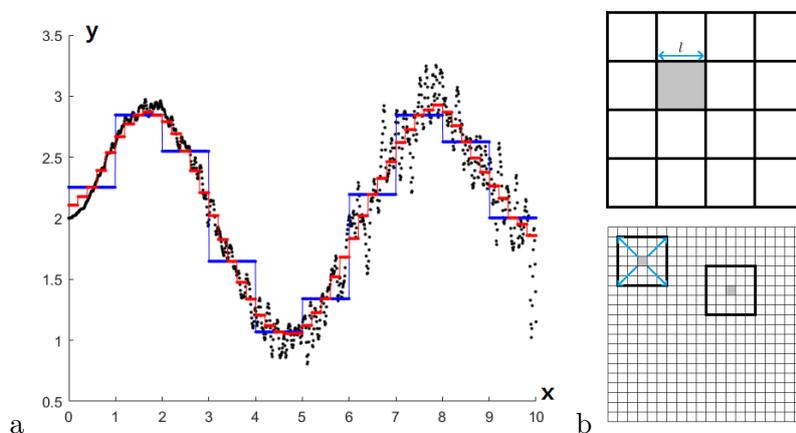


Figure 4.26: a) Exemple de calcul de moyennes locales sur un signal 1D de la forme d'un sinus auquel on a ajouté un bruit aléatoire dont l'amplitude est fonction de la variable x , en bleu sur des fenêtres de tailles caractéristiques L , en rouge sur des plus petites fenêtrées associées à des zones de taille caractéristique l . b) Schéma illustrant le calcul de moyennes de quantités en 2 dimensions.

L'estimateur usuel de l'espérance est la moyenne:

$$\mathbb{E}(X) = \frac{1}{n} \sum_{i=1}^n X_i \quad (4.14)$$

Il est possible de pondérer cette moyenne avec des coefficients α_i :

$$\tilde{\mathbb{E}}(X) = \frac{\sum_{i=1}^n \alpha_i X_i}{\sum_{i=1}^n \alpha_i} \quad (4.15)$$

Dans ce cas, X représente des quantités telles que la porosité, l'épaisseur des travées osseuses, les densités cellulaires, immunoscores, etc. L'indice i sera la position d'un pixel ou d'un élément

du pavage. Par exemple, il est possible de calculer la valeur moyenne d'une quantité sur une image ou sur une carte u en pondérant par rapport à la position (i_0, j_0) :

$$\tilde{\mathbb{E}}(X) = \frac{\sum_{i,j=-n}^n \alpha_{i,j} u(i_0 + i, j_0 + j)}{\sum_{i,j=-n}^n \alpha_{i,j}} \quad (4.16)$$

où les coefficients sont:

$$\alpha_{i,j} = \exp\left(-\frac{i^2 + j^2}{2\sigma^2}\right) \quad (4.17)$$

Définis de cette manière, les $\alpha_{i,j}$ donneront un poids de moins en moins important lorsqu'on s'éloigne de (i_0, j_0) . Par changement d'indices, $u(i_0 + i, j_0 + j)$ devient $u(i_0 - i, j_0 - j)$ (ça reste équivalent lorsque les coefficients vérifient $\alpha_{-i,-j} = \alpha_{i,j}$), on retrouve la discrétisation d'un produit de convolution:

$$\tilde{\mathbb{E}}(X) = \frac{\sum_{i,j=-n}^n \alpha_{i,j} u(i_0 - i, j_0 - j)}{\sum_{i,j=-n}^n \alpha_{i,j}} \quad (4.18)$$

Une version continue de cette moyenne locale est le flou gaussien F traduisant le produit de convolution d'une fonction f quelconque avec une gaussienne φ .

Il y a un choix pour l'écart-type σ .

$$F(f, \sigma) = f * \varphi \quad \text{où} \quad \varphi(x, y) = \frac{1}{2\pi\sigma} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (4.19)$$

Dans le figure 4.27 on calcule la moyenne en utilisant le produit de convolution F . Cela nous permet d'obtenir une version "continue" en comparaison avec la figure 4.26 dans laquelle la moyenne est obtenue sur des fenêtres.

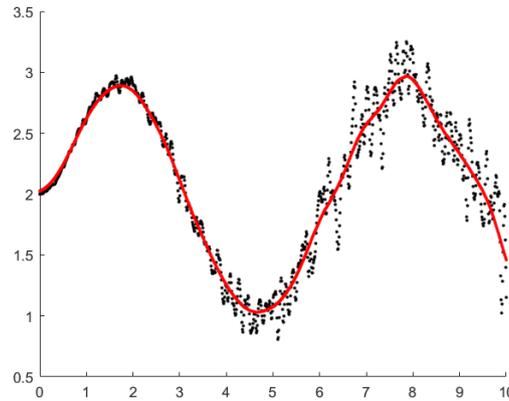


Figure 4.27: Exemple de calcul de moyenne locale sur un signal 1D par convolution.

En poursuivant cette logique, on peut définir les covariance $\tilde{C}ov$, variance $\tilde{V}ar$ et coefficient de corrélation \tilde{r} , pour deux données X et Y par :

$$\tilde{C}ov(X, Y) = \tilde{\mathbb{E}}((X - \tilde{\mathbb{E}}(X))(Y - \tilde{\mathbb{E}}(Y))) = \frac{\sum_{i=1}^n \alpha_i \beta_i (X - \tilde{\mathbb{E}}(X))(Y - \tilde{\mathbb{E}}(Y))}{\sum_{i=1}^n \alpha_i \beta_i} \quad (4.20)$$

$$\tilde{V}ar(X) = \tilde{C}ov(X, X) = \frac{\sum_{i=1}^n \alpha_i^2 (X - \tilde{\mathbb{E}}(X))^2}{\sum_{i=1}^n \alpha_i^2} \quad (4.21)$$

$$\tilde{r}(X, Y) = \frac{\tilde{C}ov(X, Y)}{\sqrt{\tilde{V}ar(X)}\sqrt{\tilde{V}ar(Y)}} \quad (4.22)$$

Remarquons qu'avec l'utilisation de poids, on n'a plus l'égalité $\tilde{\mathbb{E}}((X - \tilde{\mathbb{E}}(X))(Y - \tilde{\mathbb{E}}(Y))) = \tilde{\mathbb{E}}(XY) - \tilde{\mathbb{E}}(X)\tilde{\mathbb{E}}(Y)$.

Lorsque le coefficient de corrélation r est global, c'est un nombre, lorsqu'il est local \tilde{r} , il est de la même dimension que X et Y . Dans notre cas X et Y sont des matrices de cartes de quantités telles que la porosité et la densité cellulaire et \tilde{r} est également une matrice que l'on pourra représenter en carte. Ce coefficient de corrélation est assimilable à un coefficient de Pearson local.

Afin de rendre les résultats plus "lisses", le produit de convolution 4.19 peut être utilisé pour deux fonctions f et g quelconques et on a :

$$\tilde{Cov}(f, g) = F((f - F(f, \sigma))(g - F(g, \sigma)), \frac{\sigma}{\sqrt{2}}) \quad (4.23)$$

$$\tilde{Var}(f) = F((f - F(f, \sigma))^2, \frac{\sigma}{\sqrt{2}}) \quad (4.24)$$

Le $\frac{\sigma}{\sqrt{2}}$ vient du fait que dans les formules de la covariance et variance, la dernière convolution se fait par $\frac{\varphi^2}{\|\varphi^2\|_{L_1}}$.

Dans la figure 4.28, on montre un exemple de variance locale. On observe une variance qui augmente en fonction du bruit et de la "pente moyenne" des valeurs.

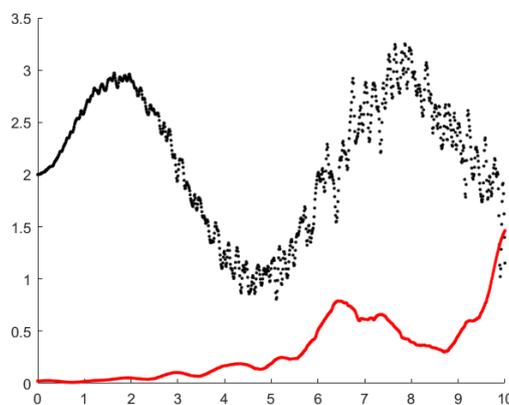


Figure 4.28: Exemple de variance locale (en rouge) sur un signal 1D (en noir) par convolution.

La figure 4.29 illustre le coefficient de corrélation local, figure 4.29b, calculé avec des convolutions F entre deux quantités illustrées figure 4.29a. La première quantité X en bleu, est un vecteur aléatoire de réalisations de lois uniformes $\mathcal{U}([0, 1])$. La deuxième quantité Y en rouge, est sur la première moitié proportionnel à X avec un coefficient 0.9, puis suit une loi uniforme indépendante de X .

Théoriquement, le profil du coefficient de corrélation locale devrait être de 1 sur la première partie, et de 0 sur la seconde. Si $\tilde{\rho} = 1$ est bien atteint dans la première partie, du bruit apparaît dans la zone $\tilde{\rho} \approx 0$. Par ailleurs, la zone de transition entre $\tilde{\rho} = 1$ et $\tilde{\rho} \approx 0$ est de taille non nulle et peut dépendre de σ . Si σ est trop petit le coefficient de corrélation local détecte bien le changement de corrélation mais la courbe serait bruitée, en revanche si σ est grand la transition entre les zones corrélées et indépendantes est de taille plus importante.

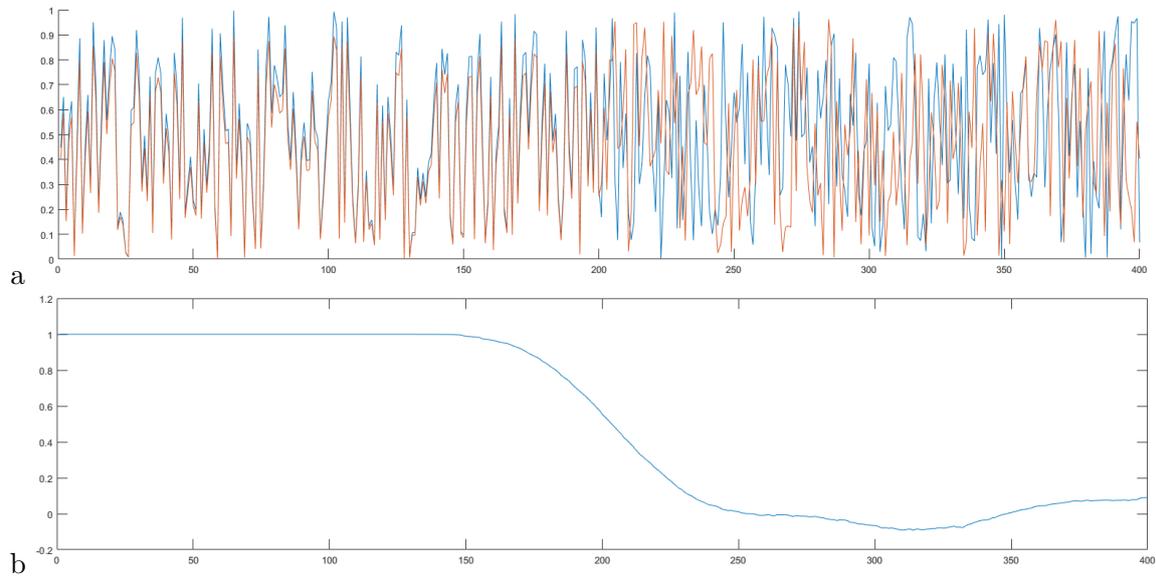


Figure 4.29: a) Signal 1D de deux jeux de données proportionnels sur la première moitié du signal, indépendants sur la deuxième moitié, b) calcul du coefficient de corrélation local associé aux signaux bleu et rouge.

Nous souhaitons maintenant appliquer cette image aux cartes de porosité et de densité cellulaires. Dans la figure 4.30, nous pouvons voir un exemple de localisation de quantités telles que la porosité et des densités cellulaires calculées sur des fenêtres de largeur 800 pixels (soit $l = 370\mu m$).

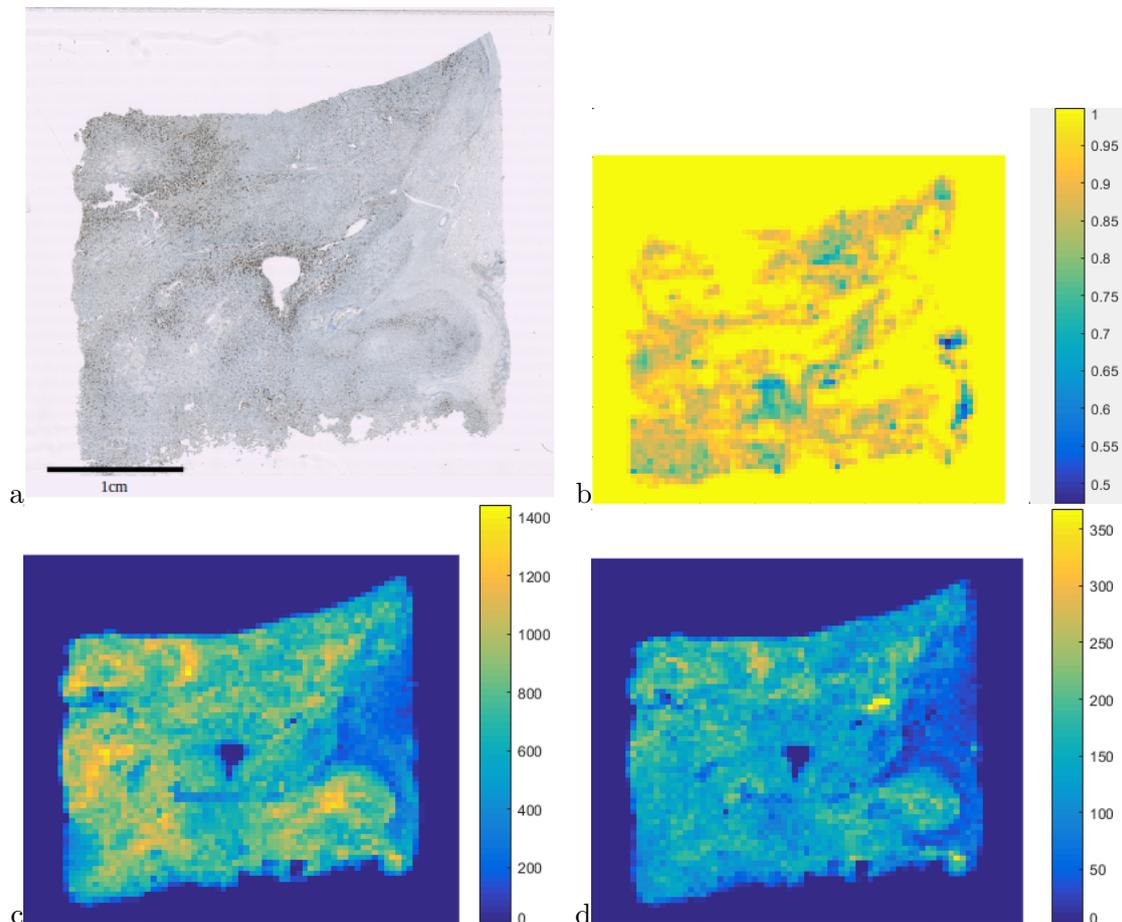


Figure 4.30: Cartes de quantités sur une coupe immuno-histologique marquée au CD68, a) Image initiale, b) porosité, et c) densité de cellules totales, d) densité de cellules marquées.

Les quantités telles que la porosité et la densité cellulaire pourraient être rendues continues sur les images de coupes histologiques et immunohistologiques, mais étant donnée la taille des images, nous faisons le calcul de corrélation locale sur des cartes avec des petites fenêtres de $80px \times 80px$, dont le contenu représente la valeur de la variable sur une fenêtre environnante de $800px \times 800px$. Une deuxième approche consiste à effectuer un flou gaussien sur les données des cartes $80px \times 80px$ précédemment définies.

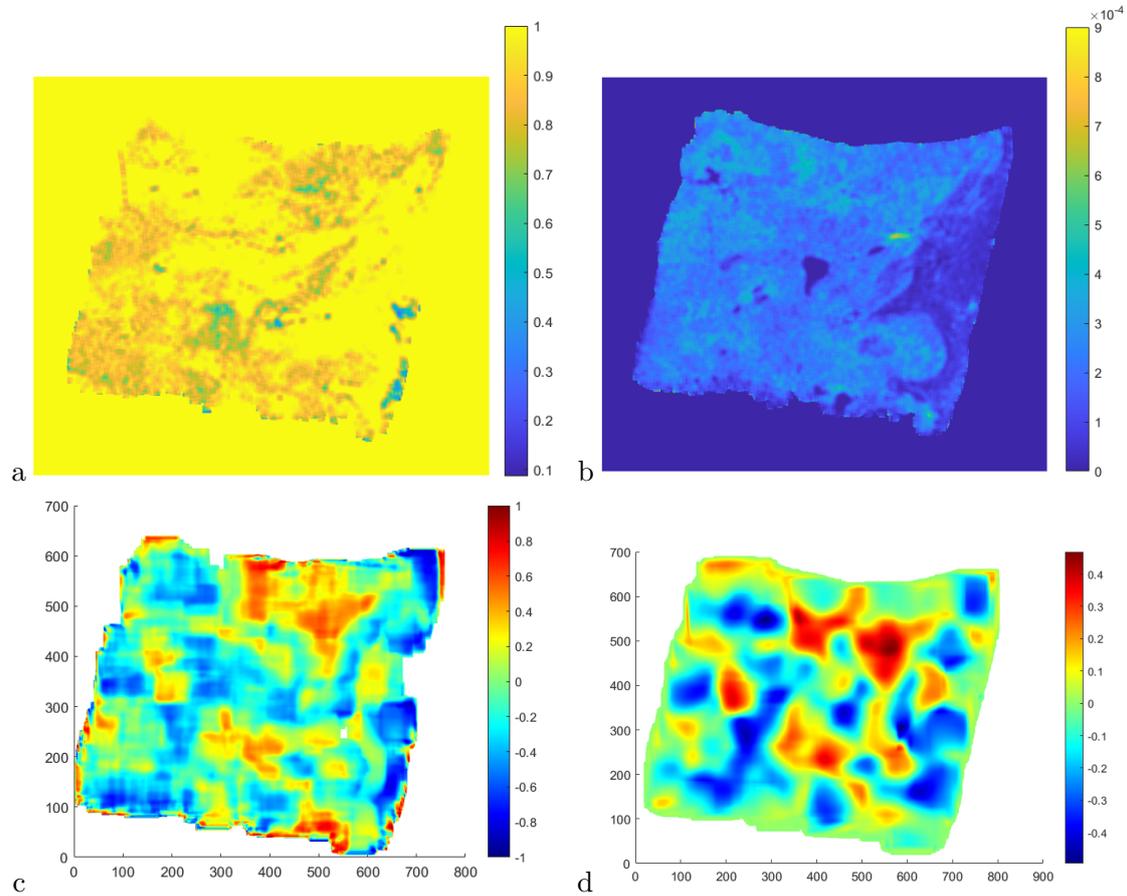


Figure 4.31: Cartes avec zones de largeurs 80 pixels, a) porosité, b) densité de cellules marquées. Cartes de corrélation locale entre la porosité et la densité de cellules marquées (CD68), c) calculée sur de petites fenêtres $80px \times 80px$, d) calculé par convolution avec une gaussienne, $\sigma = 20$.

La figure 4.31 montre les cartes de corrélations locales calculées sur des fenêtres de taille $80px \times 80px$ et avec produit de convolution avec une gaussienne. La figure 4.32 montre la répartition de la porosité et de la densité de cellules marquées dans des zones où \tilde{r} est grand et proche de 0 pour illustration.

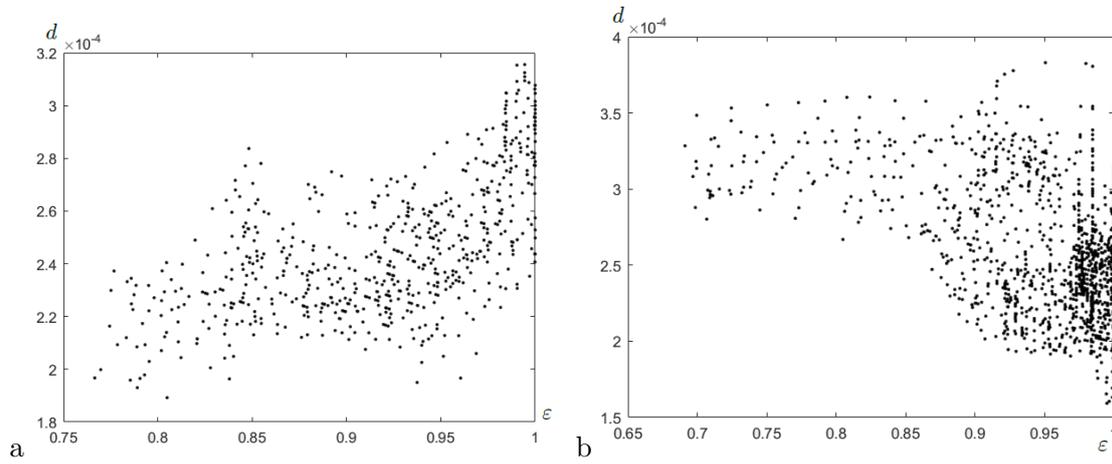


Figure 4.32: Repartition de la porosité (abscisses) et de la densité de cellules marquées (ordonnées) dans un voisinage où **a)** $\tilde{\rho}$ est grand, **b)** $\tilde{\rho}$ est petit.

La figure 4.33 montre l'évolution de \tilde{r} en fonction de σ sur des positions différentes de l'image. On peut observer la dépendance des coefficients de corrélation au σ lorsqu'on effectue les produits de convolution. Aux grands σ , on retrouve la valeur du coefficient de corrélation global calculé sur l'ensemble de l'image. Dans la suite, le choix de σ est de 20 pour préserver la capture de l'hétérogénéité du milieu. Une proposition pour calculer de la même manière un coefficient de Spearman local est proposé en annexe B.

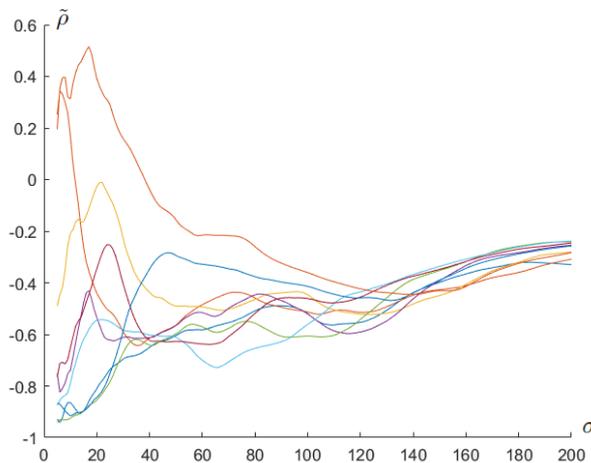


Figure 4.33: Évolution du coefficient de corrélation local en fonction de σ sur des positions différentes de l'image.

L'établissement du coefficient de corrélation local reste délicat car le choix de la taille associée à l'estimation de cette corrélation peut faire varier le signe de la corrélation de façon significative. Cela montre qu'il y a des quantités qui peuvent présenter une corrélation négative sur une certaine échelle, par exemple sur des oscillations qui varient en opposition, puis redevenir positive sur une plus grande échelle car la tendance est de varier de la même façon. L'étude de la corrélation locale dépend de l'échelle à laquelle on veut étudier des phénomènes. Une étude multi-échelle de la corrélation peut être envisagée dans le cas de l'ostéosarcome en faisant varier le choix du rayon de manière continue.

4.6.3 P-value et p-value locale

La p-value est un outil couramment utilisé dans beaucoup de domaines scientifiques lorsque les résultats obtenus sont issues de tests statistiques et que le résultat peut fluctuer d'une expérience à l'autre. La p-value mesure la probabilité que le résultat soit un faux positif, lorsque la celle-ci est sous un seuil de 0.05, on dit que le résultat est sûr à plus de 95%.

Cependant, la p-value est discutée aussi bien sur le choix arbitraire du seuil à 95% que sur la manière dont elle est établie. Dans la pratique, le calcul de la probabilité de faux positifs demanderait de choisir une loi de probabilité sur un espace d'expériences associées à l'hypothèse non nulle, cette loi est souvent purement hypothétique. La p-value est souvent remplacée par les faux négatifs, selon l'hypothèse que les deux valeurs seraient proches, ce qui est discutable.

Matlab fournit une p-value avec le coefficient de corrélation lorsqu'on utilise la fonction `corrcoef`. L'hypothèse nulle considérée dans cette p-value est : Si le r expérimental est positif, alors l'hypothèse est que la corrélation est positive, inversement si r est négatif alors l'hypothèse est que la corrélation est négative.

Par exemple, si le r obtenu est de 0,4, la p-value est la probabilité qu'en expérience de coefficient de corrélation est 0,4 aurait pu fournir un résultat négatif. Par exemple, si on génère n réalisations de deux variables aléatoires X et Y dont la corrélation est de $r_{\text{théorique}}$, et qu'on calcule le coefficient de corrélation $r(X_1, \dots, X_n, Y_1, \dots, Y_n)$ basé sur les n réalisations, alors le coefficient ne sera pas forcément exactement $r_{\text{théorique}}$ mais va suivre une densité de probabilité dépendant de $r_{\text{théorique}}$ et de n .

Dans la pratique, on ne connaît pas $r_{\text{théorique}}$, on suppose alors que $r(X_1, \dots, X_N, Y_1, \dots, Y_N)$ est le $r_{\text{théorique}}$, et on calcule la probabilité que des expériences de coefficient r donne un résultat de signe opposé.

Le calcul explicite de cette *p-value* en ayant la fonction de Matlab `tcdf` (Student t-cumulative distribution function) est, où N est le nombre de réalisations:

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} \quad (4.25)$$

$$p = 1 - \text{tcdf}(|t|, N-2) \quad (4.26)$$

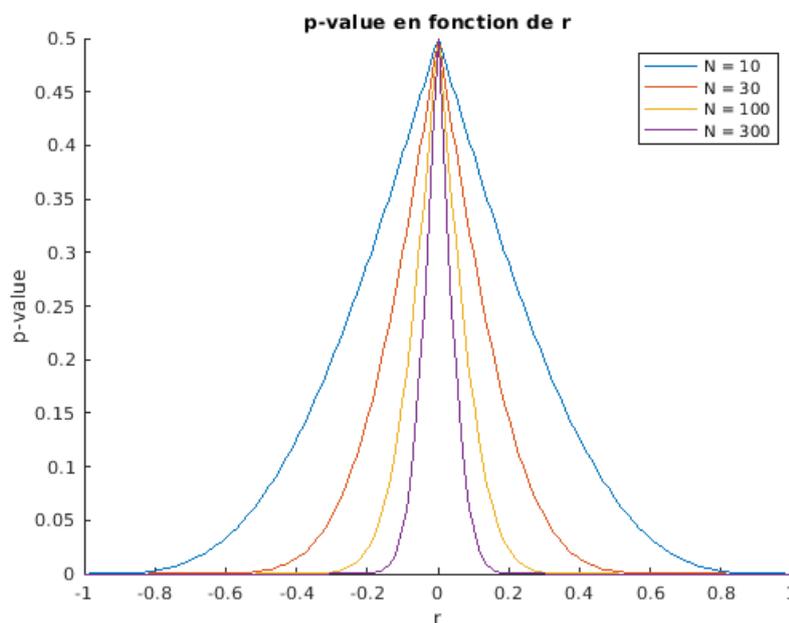


Figure 4.34: Courbes de p-value en fonction de r et N , où r est le coefficient de corrélation et N est le nombre de réalisations.

Globalement, dans la figure 4.34 on observe une diminution de la p-value lorsque N augmente et en r fixé, ce qui est tout à fait normal : lorsqu'on augmente la taille des données traitées, on augmente la significativité du résultat. Cependant, il y a une persistance de p-value élevée lorsque r est proche de 0. Si, suite à une expérience, nous obtenons un coefficient de corrélation de $r = 0,01$, ce que mesure cette p-value selon cette définition est la probabilité que la corrélation aurait pu être négative, d'où une valeur de p proche de 0,5.

Nous comprenons ici que pour une valeur de $r = 0,01$, la question pertinente n'est pas le signe

de r , mais plutôt si le résultat de non corrélation est significatif. Par exemple, quelle serait la probabilité que la corrélation se situe dans $[-0.1, 0.1]$?

Nous pouvons alors proposer une autre définition de la p-value : étant donné $r_{\text{théorique}}$, quel est la probabilité que $r_{\text{expérimental}}$ se situe dans l'intervalle $[r_{\text{théorique}} - 0, 1, r_{\text{théorique}} + 0, 1]$. Cependant, à notre connaissance, il n'y a pas de formule explicite qui donnerait cette probabilité. Nous allons approcher cette p-value par une expérience.

Soient, $X, Y \sim \mathcal{N}(0, 1)$ iid,

On peut recréer toutes les situations de corrélation en s'appuyant sur les variables X et $Z = \cos(\theta)X + \sin(\theta)Y$, il suffit que θ soit dans $[0, \pi]$ pour parcourir continument toutes les valeurs de r dans $[-1, 1]$.

$$\text{Cov}(X, Z) = \cos(\theta)\text{Cov}(X, X) + \sin(\theta)\text{Cov}(X, Y) = \cos(\theta) \quad (4.27)$$

$$\text{Var}(X) = 1 \quad (4.28)$$

$$\text{Var}(Z) = \cos^2(\theta)\text{Var}(X) + 2\cos(\theta)\sin(\theta)\text{Cov}(X, Y) + \sin^2(\theta)\text{Var}(Y) = \cos^2(\theta) + \sin^2(\theta) = 1 \quad (4.29)$$

Donc :

$$r_{\text{théorique}} = \frac{\text{Cov}(X, Z)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Z)}} = \cos(\theta) \quad (4.30)$$

Etant donné que $\theta \in [0, \pi] \Rightarrow \sin(\theta) \geq 0$, on a $\sin(\theta) = \sqrt{1 - r^2}$.

On peut donc générer deux variables aléatoires de corrélation r avec X et $Z = rX + \sqrt{1 - r^2}Y$. La p-value s'approxime en répétant un grand nombre de fois le calcul de $r_{\text{expérimental}}$ sur N réalisations de X et Z et en calculant la fréquence de valeurs en dehors de l'intervalle $[r - 0.1, r + 0.1]$.

Cette hypothèse s'adapte mieux aux cas de r proche de 0, mais elle est également plus exigeante lorsque $|r| > 0.1$ car la p-value sera plus grande qu'avec l'hypothèse du signe de r :

Si $|r| > 0.1$:

$$\left\{ \begin{array}{l} \text{et } r \geq 0, [r - 0.1, r + 0.1] \subset [0, 1] \Rightarrow P(r_{\text{exp}} \notin [r - 0.1, r + 0.1]) \geq P(r_{\text{exp}} \notin [0, 1]) \\ \text{et } r \leq 0, [r - 0.1, r + 0.1] \subset [-1, 0] \Rightarrow P(r_{\text{exp}} \notin [r - 0.1, r + 0.1]) \geq P(r_{\text{exp}} \notin [-1, 0]) \end{array} \right. \quad (4.31)$$

La figure 4.35 montre les courbes de p-value en fonction de r , avec test associé à l'intervalle $[r - 0.1, r + 0.1]$, pour $N = 100$, $N = 300$, et $N = 1000$. On observe que la valeur maximale de p se trouve en $r = 0$. On peut alors tracer p en $r = 0$ pour déterminer à partir de quelle valeur de N la p-value sera inférieure à 0.05 quelque soit r .

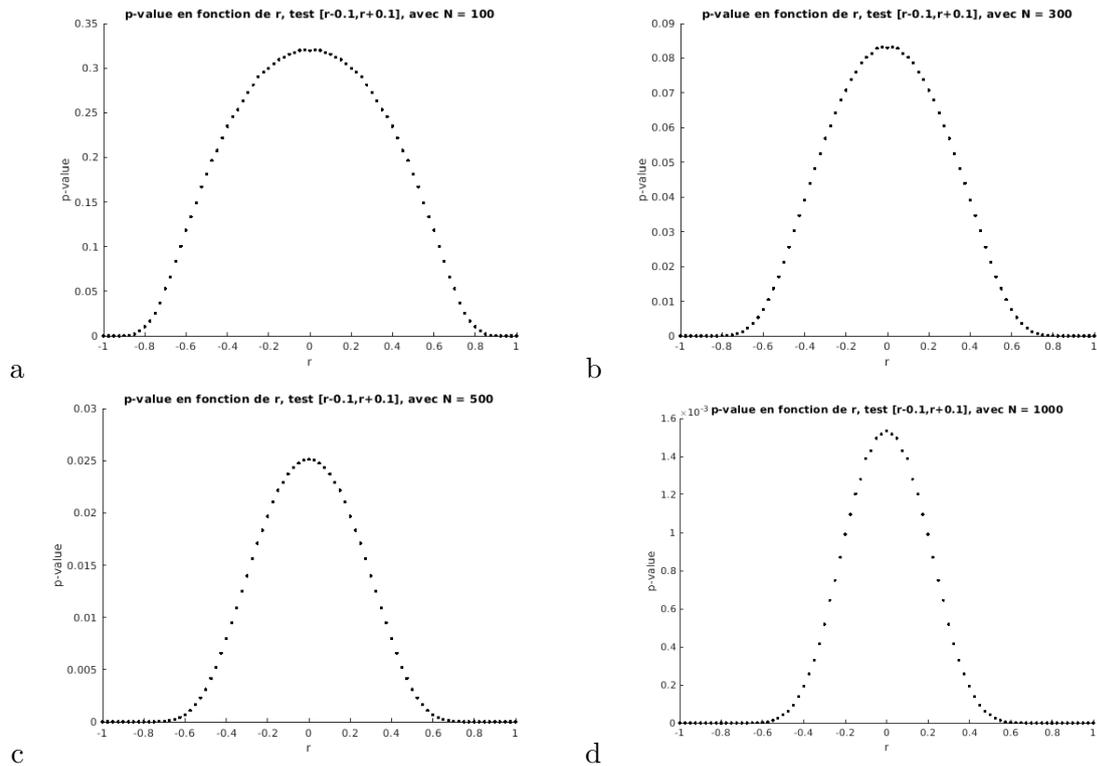


Figure 4.35: Courbes de p-value en fonction de r , avec test associé à l'intervalle $[r - 0.1, r + 0.1]$ et **a) $N = 100$, b) $N = 300$, c) $N = 500$ et d) $N = 1000$ (échelle à 10^{-3}).**

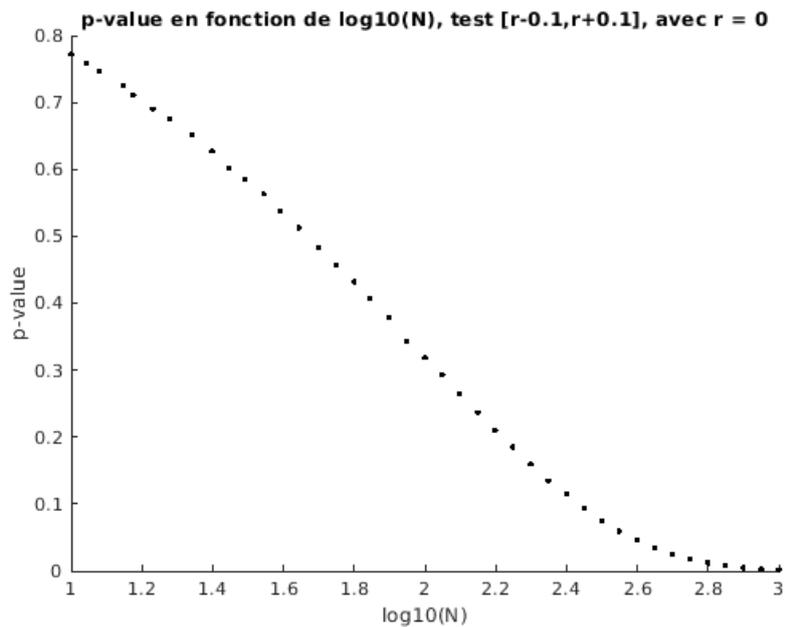


Figure 4.36: Courbes de p-value en fonction de $\log_{10}(N)$ avec test associé à l'intervalle $[r - 0.1, r + 0.1]$ et en $r = 0$.

la figure 4.36 montre que la p-value devient inférieure à 0.05 lorsque $\log_{10}(N) \approx 2.6$ (en $\log_{10}(N) = 2.6$, on y lit $p \approx 0.046$), soit environ $N = 400$. Lorsque $N \geq 400$ la p-value est inférieure à 0.05 seulement à l'extérieur d'un intervalle centré en $r = 0$.

Nous avons déjà défini la corrélation locale dans la sous partie précédente. Nous cherchons maintenant à caractériser la significativité, des cartes de corrélation locale, également de manière locale.

Il faut définir un n local, qui s'adapte à la définition de la corrélation locale.

Les expressions ((4.20)) ((4.21)) et ((4.22)) font intervenir des coefficients qui donne un poids aux valeurs en fonction de leur éloignement spatial. Posons que le poids central est de 1, ainsi le n local sera défini par :

$$n_{loc} = \frac{1}{\alpha_{0,0}} \sum_{i,j=-n}^n \alpha_{i,j} \quad (4.32)$$

Avec en général :

$$\alpha_{i,j} = \exp\left(-\frac{1}{2\sigma^2}((i - i_0)^2 + (j - j_0)^2)\right) \quad (4.33)$$

Le facteur de pertinence $p_{i,j}$ n'est pas pris en compte dans la normalisation des poids parce que si la valeur centrale est peu pertinente, son poids doit être inférieur à 1. Ainsi, la p-value locale sera importante si nous sommes dans une zone de la lame peu pertinente.

$$n_{loc} = \frac{1}{\alpha_{0,0}} \sum_{i,j=-n}^n \alpha_{i,j} p_{i,j} \quad (4.34)$$

En général, le facteur de pertinence est le produit entre le pourcentage de pixels présents dans la lame de tissu et la porosité. Appliqué avec $\sigma = 20$, une porosité de 0.6 (soit $p_{i,j} = 0.6$ dans la lame) nous aurons $n \approx 1381$. Et si nous sommes dans un coin de la lame formant un angle de 90° , n descend à $n \approx 345$.

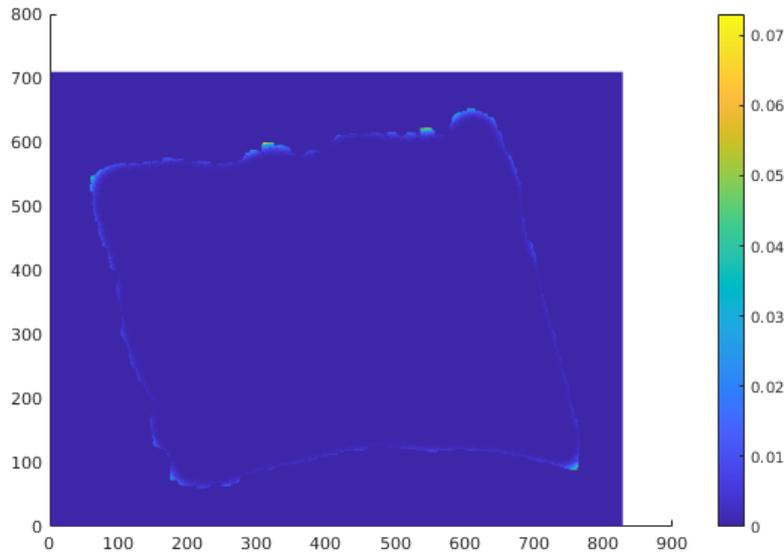


Figure 4.37: Carte de p-value locale

La figure 4.37 montre un exemple de carte de p-value locale appliquée sur une lame et nous voyons que les zones non significatives se limitent au bord du tissu.

Pour conclure ce chapitre, nous avons montré comment calculer de manière rigoureuse différentes métriques géométriques dans l'ostéosarcome que sont la porosité, le périmètre spécifique, la largeur des travées, la densité cellulaire. Ces calculs font intervenir des algorithmes de calculs de facteur de pertinence et éventuellement de recalage. Dans tous les cas, en suivant l'approche de Durlofsky [47] utilisée pour l'étude du transport en milieux poreux, ces quantités sont données sur un pavage de l'image donnant des cartes facilement interprétables. A partir du calcul de ces métriques, un calcul innovant de coefficient de corrélation local a été proposé afin de calculer la distribution relative des populations cellulaires discrètes relativement à la phase continue osseuse et ce, en préservant des informations sur l'hétérogénéité du milieu.

Dans la suite, l'exploration de ces cartes de corrélations locales guideront la recherche de marqueurs de la réponse au traitement dans l'ostéosarcome.

Chapter 5

Application des algorithmes sur une cohorte de patients

Contents

5.1	Correlations globales	93
5.2	Patients mauvais répondeurs: patient 2	94
5.3	Patient bon répondeur: patient 7	97
5.4	Conclusion	99

Dans ce chapitre, les bons et mauvais répondeurs ont été définis via le score Huvos&Rosen [60] par l'anatomopathologiste sur un ensemble de coupes histologiques. Dans la suite, les coupes immuno-histologiques étudiées peuvent cependant présenter une hétérogénéité dans la réponse au traitement avec une distribution spatiale complexe de zones de bonnes et mauvaises réponses. Les patients 1 à 3 sont des mauvais répondeurs (MR) au traitement alors que les patients 4 à 8 sont bons répondeurs (BR). Nous nous attacherons d'abord à explorer les corrélations globales, puis analyserons les résultats des corrélations locales pour un patient mauvais répondeur puis un bon répondeur. L'ensemble des résultats des corrélations locales est disponible en annexe C. L'objectif de cette approche est d'identifier des marqueurs de bonnes réponses au traitement prenant en compte l'hétérogénéité spatiale de ces tumeurs en terme de diversité de populations cellulaires immunitaires, de formation ou resorbtion osseuse et de traitement.

5.1 Correlations globales

Dans cette section, nous nous sommes intéressés dans un premier temps aux coefficients de corrélation globaux r . Le tableau 5.1 indique les résultats pour les 8 patients de la cohorte et ceci pour trois marquages CD8, CD163 et CD68. Le marquage CD68 est étudié de deux manières, soit sur l'ensemble des cellules marquées CD68t, soit uniquement sur les grosses cellules dont le diamètre est supérieur à $20\mu m$ CD68g. Ces cellules correspondent aux cellules polynuclées (figure 5.1), qui sont des ostéoclastes matures dont la présence en abondance est de mauvais pronostic pour les patients [11]. Lorsque l'amplitude $|r|$ du coefficient est supérieure à 0.1, la p-value est très inférieure à 0.05. Pour les amplitudes plus faible cette p-value augmente fortement et les résultats sont peu significatifs. L'analyse de ce tableau montre qu'il est difficile de dégager une tendance sur les coefficients de corrélation globaux pour les MR. En revanche pour les BR, les macrophages de bon pronostic CD163 [16] et les CD68t sont dans la majorité corrélé négativement avec la porosité avec certes une amplitude limité. Notons que pour les BR, les r_{CD68t} et r_{CD68g} évoluent différemment alors que pour les patients 6 à 8 CD163 et CD68t ont une tendance similaire. Ce résultat est potentiellement dû à l'existence de cellules multi-potentes marquées CD68 et CD163 simultanément [11]. Ces cellules sont principalement des petites cellules qui n'apparaissent pas dans la population CD68g.

Au vu des limites des résultats de corrélations globales, nous souhaitons explorer les résultats

de corrélations locales comme explicitées dans le chapitre précédent.

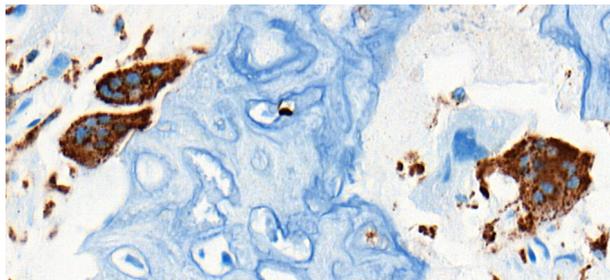


Figure 5.1: Ostéoclastes correspondant aux trois grosses cellules polynucléées de l'image. Marquage CD68.

Coefficient de corrélation global entre porosité/ densité cellulaire	CD8	CD163	CD68 t	CD68 g
Patient 1	-0.007	-0.116	-0.207	0.115
Patient 2	-0.038	-0.117	-0.022	-0.088
Patient 3	-0.462	-0.073	-0.117	-0.078
Patient 4	-0.108	-0.222	-0.092	-0.033
Patient 5	-0.267	-0.311	-0.113	-0.059
Patient 6	-0.009	-0.096	-0.332	-0.249
Patient 7	-0.090	-0.396	-0.436	0.000
Patient 8	-0.146	-0.408	-0.458	-0.180

Table 5.1: Coefficients de corrélation globaux entre porosité et densité cellulaire des cellules marquées, pour le CD8, CD163, CD68 : t = toutes cellules marquées, g = grandes cellules marquées. Patients 1 à 3 mauvais répondeurs et patients 4 à 8 bon répondeurs.

5.2 Patients mauvais répondeurs: patient 2

Dans la suite, les coupes immunohistologiques figure 5.2, les cartes de porosité 5.3, les cartes de densité cellulaires 5.4 et les cartes de corrélations locales 5.5 sont illustrées pour le patient 2, mauvais répondeur à la chimiothérapie.

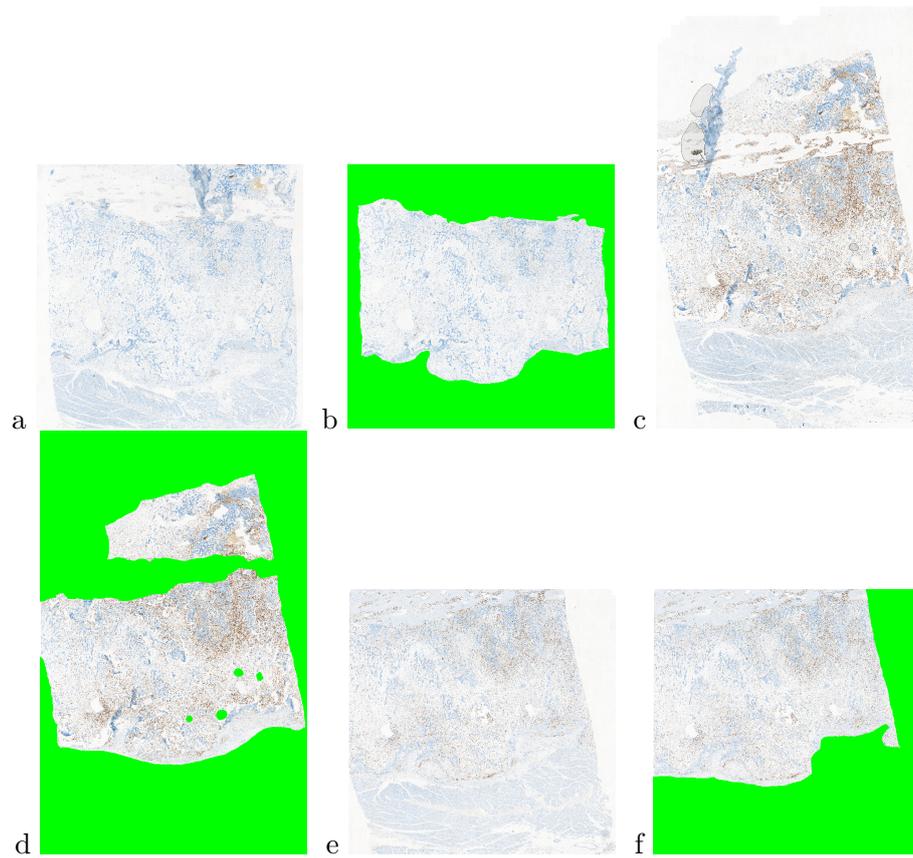


Figure 5.2: Coupes immuno-histologiques du patient 2, **a)** marquage CD8, **b)** masque CD8, **b)** marquage CD163, **d)** masque CD163, **e)** marquage CD68, **f)** masque CD68.

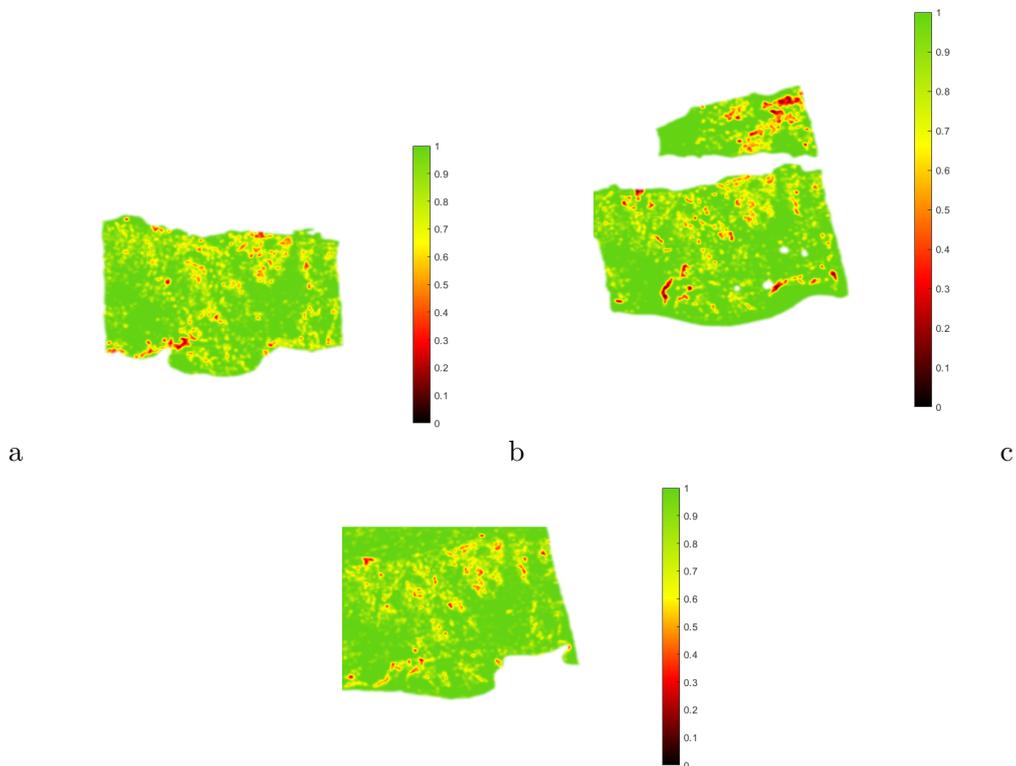


Figure 5.3: Cartes de porosité **a)** CD8, **b)** CD163, **c)** CD68

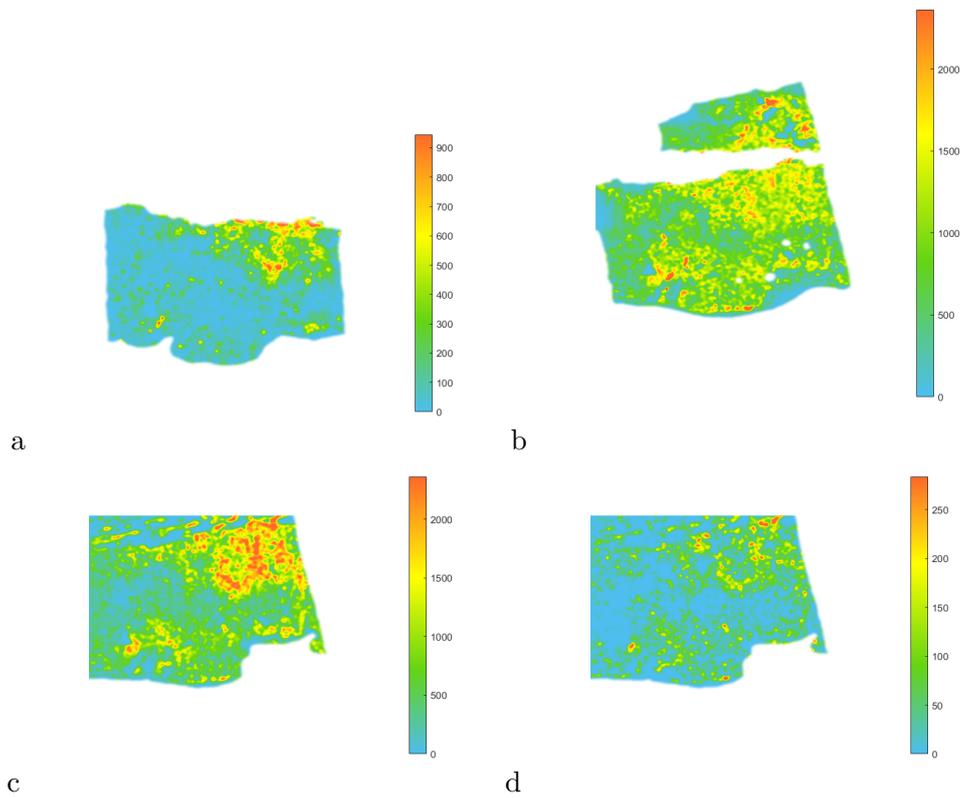


Figure 5.4: Cartes de densité cellulaire **a)** CD8, **a)** CD163, **c)** CD68 cellules totales, **d)** CD68 grandes cellules.

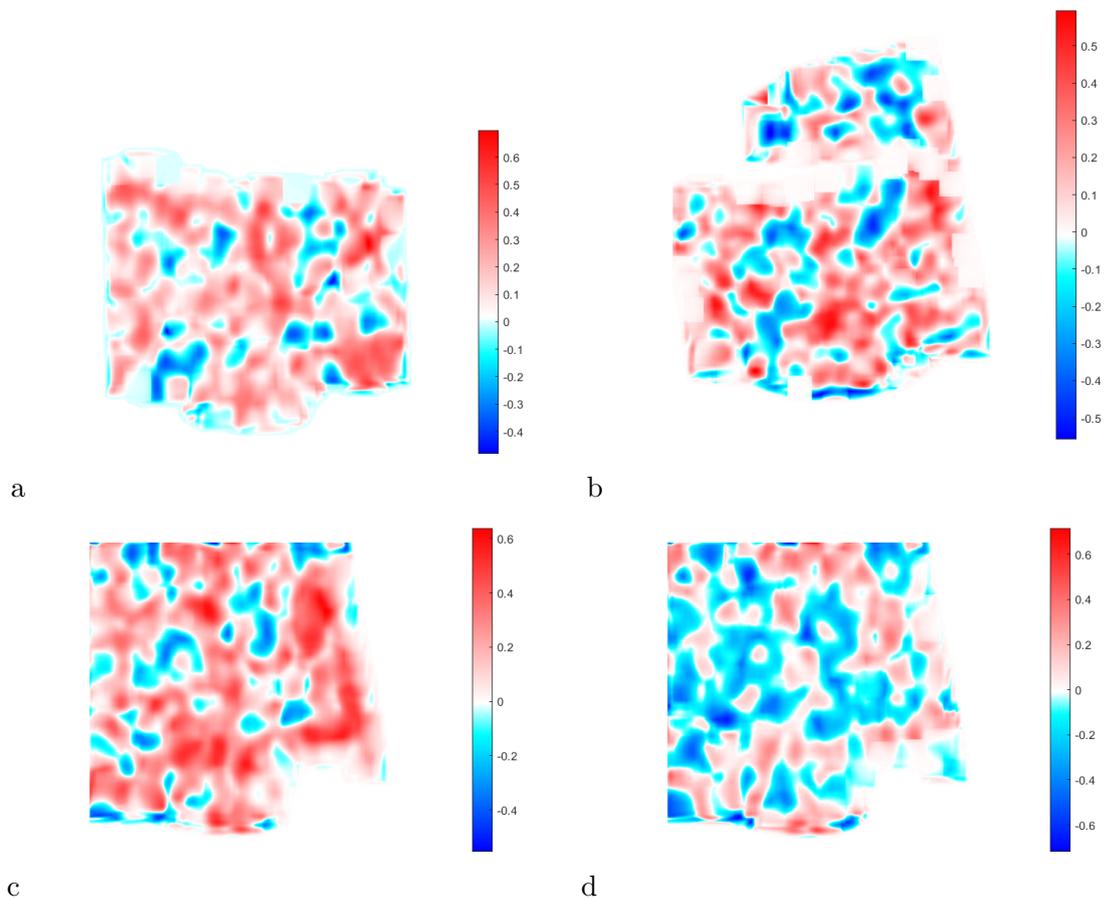


Figure 5.5: Cartes de corrélations locales **a)** CD8, **b)** CD163, **c)** CD68 cellules totales, **d)** CD68 grandes cellules.

Il est difficile d'avoir un patient représentatif des mauvais répondeurs car chacun présente des caractéristiques différentes, cependant nous pouvons déjà noter, en étudiant la carte des porosités figure 5.3 que la tumeur présente une phase osseuse étendue $\epsilon \approx 0.7$ contrairement à d'autres tumeurs lytiques ayant des porosités de l'ordre de ≈ 0.9 . L'environnement est principalement immunosuppresseur, figure 5.4a avec des densités cellulaires en lymphocyte T (CD8) majoritairement inférieures à $200mm^{-2}$. En revanche les macrophages CD163, figure 5.4b, et surtout CD68, figure 5.4c, se retrouvent en abondance. Les ostéoclastes matures, figure 5.4d, sont également en grande quantité et surtout distribués sur l'ensemble de la lame. Concernant les cartes de corrélations locales figure 5.5, on observe une très forte hétérogénéité spatiale des cartes de corrélation avec une alternance de zones de corrélation positive et négative quelque soit le marquage.

5.3 Patient bon répondeur: patient 7

Dans la suite, les coupes immunohistologiques figure 5.6, les cartes de porosité 5.7, les cartes de densité cellulaires 5.8 et les cartes de corrélations locales 5.9 sont illustrées pour le patient 7, bon répondeur à la chimiothérapie néoadjuvante.

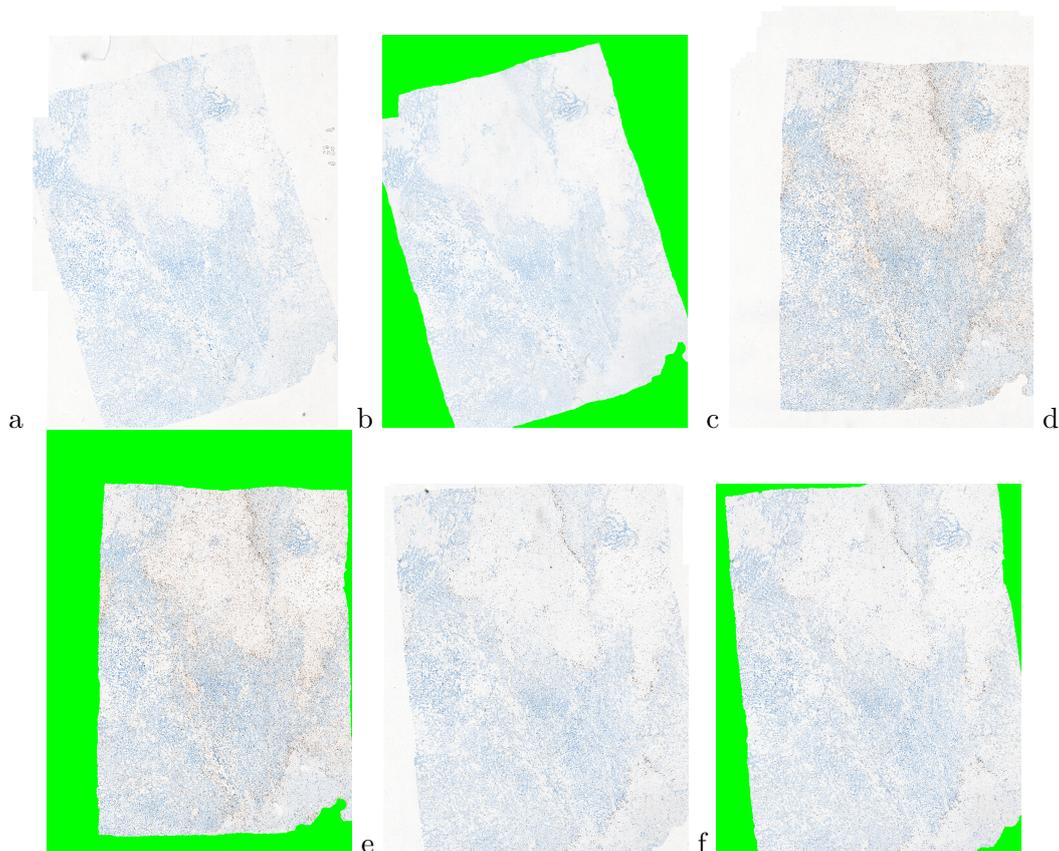


Figure 5.6: Coupe immuno-histologique du patient 21TED05, **a)** marquage CD8, **b)** CD8 masque, **c)** CD163, **d)** CD163 masque, **e)** CD68, **f)** CD68 masque.

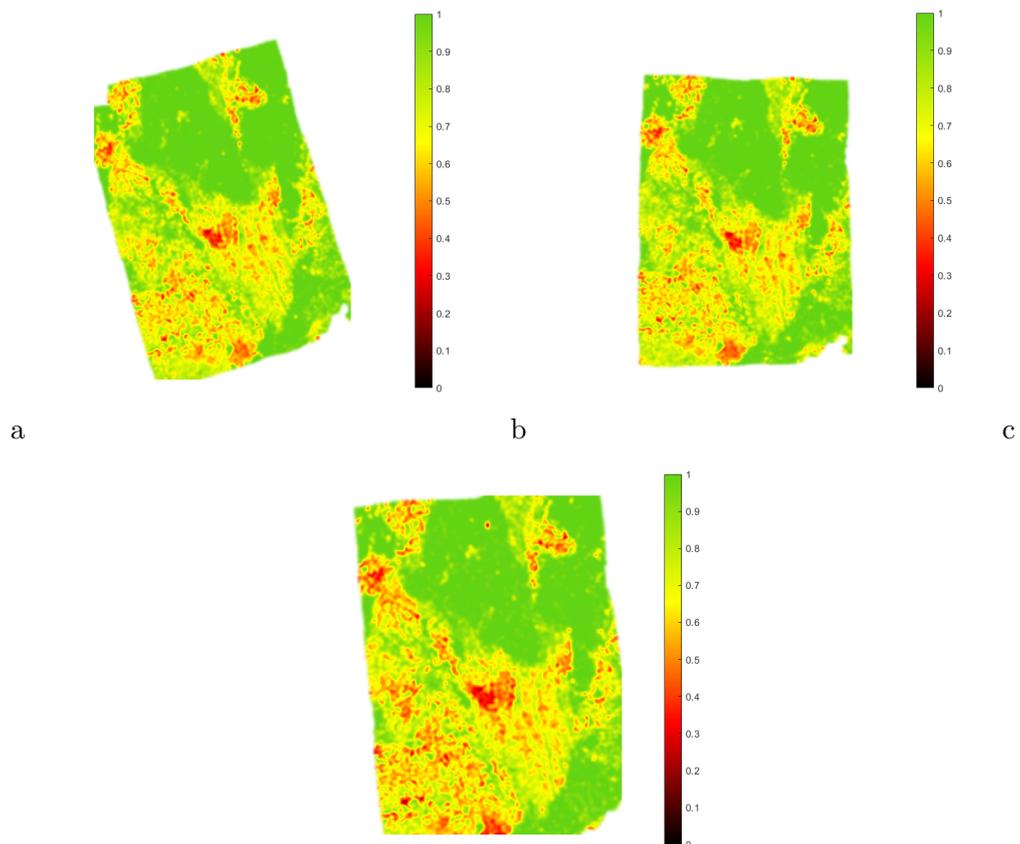


Figure 5.7: Carte de porosité a) CD8, b) CD163, c) CD68.

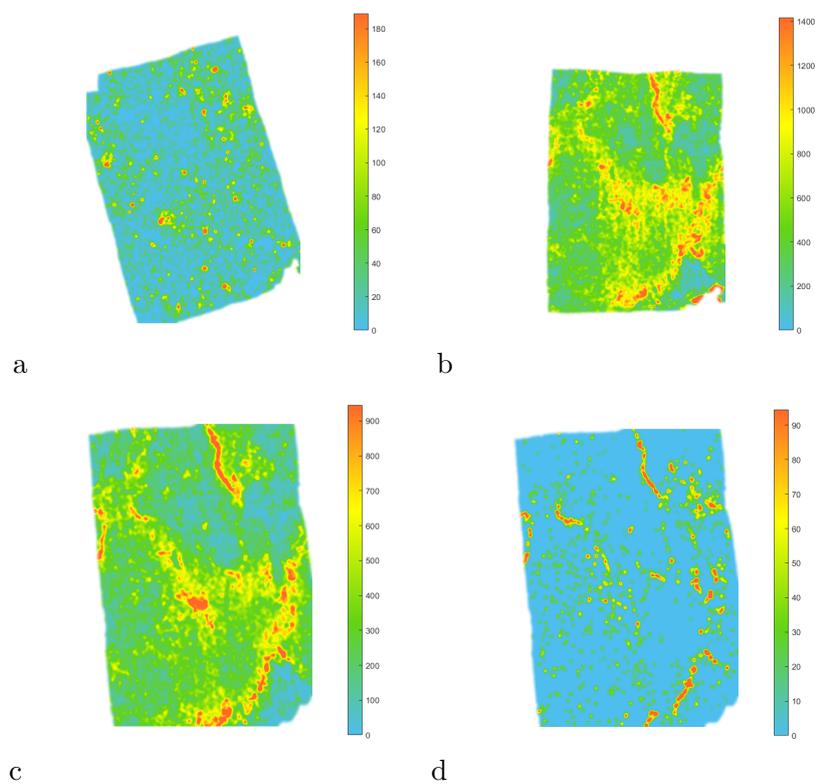


Figure 5.8: Carte de densités a) CD8, b) CD163, c) CD68 cellules totales, d) CD68 grandes cellules.

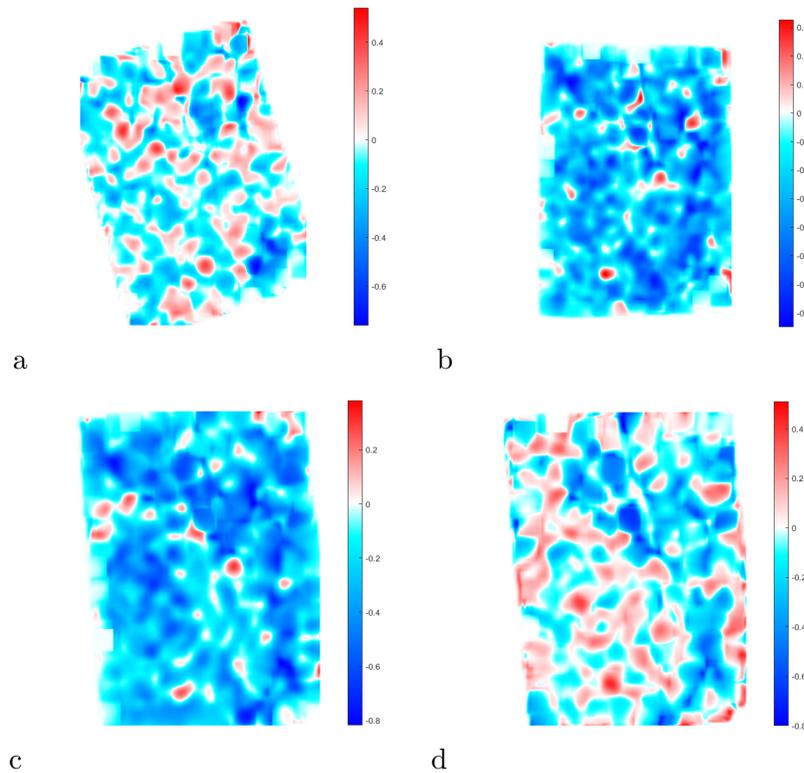


Figure 5.9: Cartes de corrélations locales **a)** CD8, **b)** CD163, **c)** CD68 cellules totales, **d)** CD68 grandes cellules.

Dans le cas des patients bons répondeurs, il y a une meilleure homogénéité entre les patients. Nous avons choisi le patients 7 pour illustrer le type de résultats obtenus. Sur la carte des porosités figure 5.7, nous pouvons observer que la tumeur présente une phase osseuse étendue $\epsilon \approx 0.3 - 0.7$ avec des zone de forte densité osseuse. L'environnement reste principalement immunosuppresseur, figure 5.8a avec des densités cellulaires en lymphocyte T (CD8) majoritairement inférieures à $100mm^{-2}$. En revanche les macrophages CD163 se retrouvent en abondance, figure 5.8b, et les macrophages CD68 en moindre quantité que chez les patients mauvais répondeurs, figure 5.8c. Les ostéoclastes matures, figure 5.8d, sont plus rares et localisés dans l'espace. Il s'agit probablement d'ostéoclastes recrutés pour résorber la matrice tumorale osseuse générée.

Concernant les cartes de corrélations locales figure 5.9, on observe une très forte hétérogénéité spatiale des cartes de corrélation pour le CD8 et le CD68g. Ces résultats sont à valider car les densité cellulaires sont faibles pour ces populations cellulaires. En revanche, les cartes pour pour les coefficients de corrélations sur CD163 et CD68t montrent une tendance principalement négative, indiquant que ces populations cellulaires augmentent lorsque la surface de phase osseuse augmente (tendance inverse à la porosité). Cette caractéristique se retrouve sur l'ensemble des patients bons répondeurs et semble ainsi traduire une caractéristique de la bonne réponse au traitement et un potentiel marqueur de celle-ci.

5.4 Conclusion

En conclusion de ce chapitre, nous avons pu identifier des premières tendances sur une cohorte de 8 patients incluant trois MR et cinq BR. Dans l'ensemble des patients, on retrouve les résultats qualitatifs de la littérature [16] indiquant que l'ostéosarcome est immunosuppresseur (CD8 de faible densité) quelque soit la réponse au traitement. Les macrophages CD68 se retrouvent en plus grande quantité chez les MR que chez les BR, les CD163 sont en abondance chez tous les patients. Les grosses cellules marquées CD68 sont en plus grande abondance chez les MR que pour les BR et distribuées sur l'ensemble de la lame dans le premier cas et localisées dans certaines régions dans le second. Ce indique probablement que dans le cas des BR, des ostéoclastes sont recrutés pour résorber la matrice osseuse formée par la tumeur. Les cartes

de corrélation locales sont très hétérogènes quelque soit le marquage chez les MR avec aucune tendance identifiée. Chez les BR, ces cartes sont également spatialement hétérogènes pour les lymphocytes CD8 et les grandes cellules CD68, un résultat à modérer du fait de la faible densité cellulaire de ces populations. En revanche chez les BR, les cartes de corrélation locales CD163 et CD68t montrent une tendance négative avec des coefficients de corrélation de grandes amplitudes indiquant que ces populations cellulaires macrophagiques augmentent dans les zones osseuses, ce qui semble être un bon marqueur de la réponse au traitement. Ces résultats corroborent le calcul du coefficient de corrélation global. Par ailleurs, le fait que cette évolution soit observée à la fois sur le CD163 et le CD68t mais pas sur le CD68g, conforte le rôle potentiellement crucial des cellules multipotentes simultanément marquées CD68 et CD163 [11] dans la réponse au traitement, et leur intérêt dans la recherche de nouvelles cibles thérapeutiques. Ces résultats doivent bien sûr être validés sur une plus grande cohorte. A très court terme, une perspective est d'explorer plus en détail les patients 3,4, 5 et 6 qui présentent des zones de bonne et mauvaise réponse sur la même coupe afin de potentiellement dégager de nouveaux résultats en fonction de la réponse locale aux traitements.

Chapter 6

Vers l'imagerie fonctionnelle microscopique numérique

Contents

6.1 Effets mécaniques et ostéosarcome	101
6.2 Méthodologie pour l'imagerie fonctionnelle numérique	102

Dans ce chapitre, nous allons voir comment les développements réalisés dans cette thèse ont permis de développer des techniques d'imagerie fonctionnelle numérique dans le cadre d'un projet plus général incluant plusieurs collaborateurs du groupe Milieux Poreux et Biologiques de l'IMFT ainsi que le Pr Anne Gomez-Brouchet du Cancéropôle de Toulouse.

6.1 Effets mécaniques et ostéosarcome

Cette approche part du constat que l'os est un organe mécanosensible particulier, qui se remodelé en permanence en fonction des forces qui lui sont appliquées [74]. Plus précisément, les effets mécaniques stimulent l'interaction entre les ostéoblastes et les ostéoclastes (OC), responsables de la formation et de la résorption osseuses. Ainsi, les os s'adaptent constamment pour maintenir leur force [75, 76], croître [77] et cicatriser [78]. La mécano-transduction a été mise en évidence [79] comme un lien direct entre le comportement cellulaire et la mécanique de l'os [80, 81]. Le caractère mécanosensible est observé au niveau d'une cellule, où il a été démontré que l'action de différents stimuli mécaniques peut modifier les fonctions de la cellule et induire un déséquilibre temporaire ou permanent dans l'homéostasie du tissu. Au niveau de l'organisme, toutes les cellules sont alors soumises à la fois à des contraintes externes, responsables de déformations dites globales (torsion, flexion, compression) [82], un exemple étant la marche, et à des contraintes locales intrinsèques [83] (figure 6.1). Ces contraintes correspondent aux flux liés à l'angiogenèse et aux fluides interstitiels [84, 85], aux effets solides tels que les déformations liées aux phénomènes de surpression (par exemple, la prolifération dans un milieu confiné) et enfin aux interactions des cellules entre elles et avec la matrice extracellulaire.

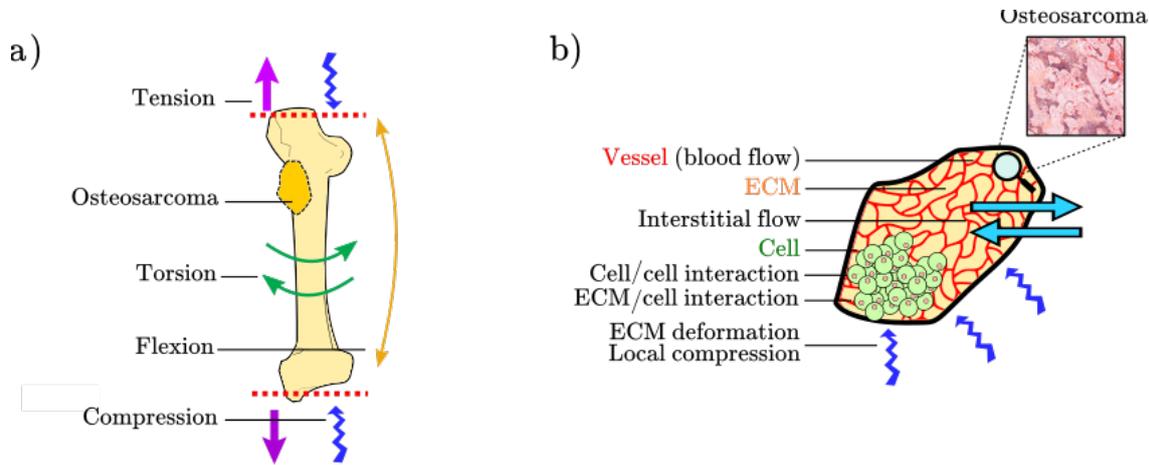


Figure 6.1: a) Effets mécaniques extrinsèques dans l'ostéosarcome, b) effets mécaniques intrinsèques dans l'ostéosarcome (d'après la thèse de doctorat de A. Moreno). Les effets mécaniques se produisent à de grandes échelles spatiales telles que la stimulation causée par l'activité physique du patient (fig. 1a) et à l'échelle locale au sein de la tumeur (fig. 1b). À l'échelle locale, les effets mécaniques identifiés comprennent, par exemple, les contraintes de compression en réponse à la prolifération cellulaire ou les changements dans le transport des fluides liés à l'angiogenèse tumorale.

Dans les cancers, tous ces phénomènes extrinsèques et intrinsèques participent à l'évolution de l'hétérogénéité spatio-temporelle des cellules tumorales et de leur microenvironnement, induisant des changements dans l'architecture et la mécanobiologie de la tumeur. Par exemple, l'angiogenèse tumorale joue un rôle crucial dans la distribution de la pression, des métabolites et de l'oxygène dans la tumeur et par conséquent dans le flux interstitiel tumoral, qui joue à son tour un rôle dans la dissémination des métastases [86]. La nature inflammatoire et/ou hypoxique de la tumeur peut également conduire à l'angiogenèse et à un risque accru de métastases par dégradation mécanique et métabolique de la matrice extracellulaire (MEC) [87]. Inversement, les auteurs de [83] mettent en évidence des liens entre différents effets biomécaniques, le microenvironnement tumoral et la progression du cancer. Il est crucial de souligner que tous ces phénomènes biologiques et mécaniques sont interconnectés et agissent les uns sur les autres de manière imbriquée à chaque échelle. En effet, il a été démontré que des effets mécaniques purement intrinsèques tels que la contrainte de cisaillement, les différences de déplacement des fluides, la compression, l'étirement, peuvent agir directement sur les cellules tumorales dans des expériences contrôlées *in vitro* [88], et modifier le phénotype et la différenciation cellulaire. Les tumeurs osseuses primaires comme l'OST, en raison de la nature des populations cellulaires affectées, montrent une grande sensibilité aux effets mécaniques locaux et globaux [82, 83].

6.2 Méthodologie pour l'imagerie fonctionnelle numérique

Dans cette partie, je vais résumer l'approche développée dans l'article en cours de soumission "Digital functional imaging: a multi-scale model to explore tumors biophysical markers from clinical histology in osteosarcoma", par A. Moreno, A. Gomez-Brouchet, A. Mancini, M. Quintard, P. Swider, P. Assemat. Dans ce travail, ma contribution principale correspond à la segmentation des images et l'établissement du pavage, ainsi que l'automatisation de ces techniques pour pouvoir étudier un grand nombre de coupes.

La méthodologie résumée dans ce chapitre est basée sur le fait que les coupes histologiques et immunohistologiques d'ostéosarcomes mettent en évidence que le tissu cancéreux d'ostéosarcome est un milieu poreux multiphasique (fluide, solide, cellules). Par ailleurs comme indiqué précédemment, le développement tumoral est multi-physique. Dans un premier temps, la phase cellulaire est considérée à part et les phases fluide et solide constituent le milieu poreux d'intérêt. Dans le cas de milieux poreux hétérogènes, les méthodes d'homogénéisation classiques [48] ou

adaptées aux tissus biologiques [89] ne sont pas directement applicables car la physique à l'échelle supérieure dépend fortement de la microstructure de l'objet. Pour surmonter cette limitation, la méthode d'homogénéisation numérique par blocs [47] est utilisée pour déterminer non seulement les équations à l'échelle supérieure qui peuvent être différentes de celles décrivant les phénomènes à l'échelle du pore (par exemple Stokes et Darcy pour le transport), mais également les propriétés de transport et mécaniques à cette échelle. Avec l'amélioration des techniques de traitement d'images, il est possible de travailler à des échelles avec des microstructures de plus en plus fines et complexes. Évidemment, une telle précision a un inconvénient : les simulations sur de très grands volumes impliquent un coût élevé. Pour résoudre ce problème, nous décidons d'ajouter à la méthode grid-block une fonctionnalité séquentielle en deux étapes [90]. Elle est caractérisée par une itération du processus à une échelle subsidiaire. L'approche séquentielle réduit le coût en mémoire et en temps de calcul tout en fournissant un encadrement précis des propriétés physiques équivalentes à différentes échelles et en préservant l'anisotropie. Une des difficultés est la prise en compte des conditions limites sur les différents blocs [91], les conditions limites réelles n'étant pas accessibles. Pour résoudre ce problème, nous avons combiné les calculs numériques avec une méthode "extend-local" où les conditions limites sont appliquées à une certaine distance de la frontière des blocs [92]. Au final des cartes de propriétés équivalentes sont obtenues. Ces cartes ressemblent à ce qu'on pourrait obtenir par des techniques *in-vivo* d'IRM de diffusion ou d'élastographie (ultra-son ou IRM) à plus grande échelle. L'avantage ici est que la distribution spatiale des cellules est connue et que nous pouvons comparer ces propriétés équivalentes à la distribution des cellules, pour par exemple étudier la résistance aux traitements. En corrélant, cartes de propriétés mécaniques équivalentes et cartes de densité cellulaire comme décrites dans le chapitre 4, nous pouvons extraire des données quantitatives sous la forme de coefficients de corrélation permettant de rechercher des mécano-biomarqueurs à partir d'images histologiques ou immunohistologiques numérisées. La stratégie générale est résumée dans la figure 6.2.

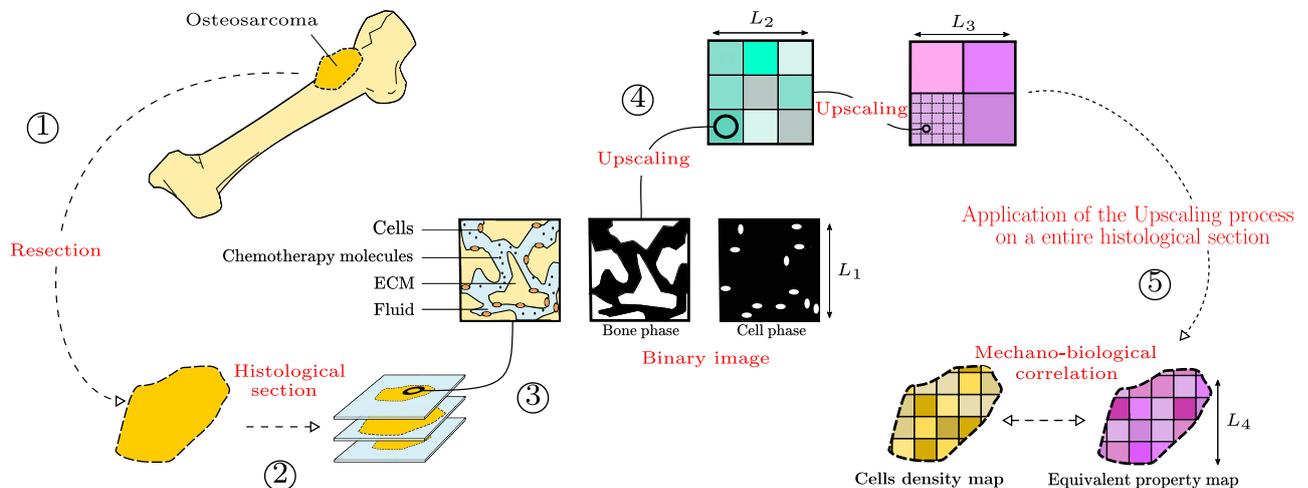


Figure 6.2: Caractérisation mécanique de l'ostéosarcome à partir de coupes histologiques. (1) L'ostéosarcome est retiré chirurgicalement du patient. (2) Une étude microscopique de la tumeur est réalisée par le pathologiste à partir de coupes histologiques d'ostéosarcomes qui sont ensuite numérisées. (3) Un traitement d'image des coupes histologiques est réalisé pour obtenir des images binaires des phases osseuses et cellulaires. (4) Un processus séquentiel de changement d'échelle Grid-Block est appliqué sur l'image de la phase osseuse. (5) Recherche d'une corrélation entre la densité cellulaire et les propriétés mécaniques équivalentes à l'échelle de la tumeur. Identification potentielle de mécano-biomarqueurs. (De Moreno et al. à soumettre)

Dans ce travail, nous avons décidé d'utiliser la méthode des éléments finis (MEF) pour résoudre les équations de transport et d'élasticité aux différentes échelles. De nombreux logiciels MEF sont disponibles, qu'ils soient gratuits (FEniCS, FreeFem++) ou commerciaux (ANSYS,

NASTRAN, COMSOL, Abaqus). Dans ce qui suit, FEniCS [93] est choisi. Cet outil nécessite une compréhension des mathématiques sous-jacentes aux FEM, mais il permet de résoudre un large spectre d'équations aux dérivées partielles et de leurs couplages par une écriture intuitive de formulations faibles. Les composants de FEniCS sont également conçus pour le traitement parallèle. Pour générer des maillages complexes, GMSH [94], un logiciel libre de génération de maillages 2D et 3D, est utilisé. La visualisation des résultats est faite via Paraview [95].

Pour résumer rapidement ce travail, l'exemple du transport dans l'ostéosarcome est proposé. A l'échelle du pore, le modèle considéré est un modèle à 1 porosité c'est-à-dire que l'on extrait la phase osseuse et on considère le reste du milieu comme du fluide. Dans ce qui suit, Ω et $\partial\Omega$ représentent respectivement les domaines et les frontières, tels que (1) $\Omega = \Omega_f \cup \Omega_s$ où Ω_f et Ω_s correspondent respectivement aux composantes fluide et solide du domaine et (2) $\partial\Omega = \partial_{int}\Omega \cup \partial_g\Omega \cup \partial_{ext}\Omega$, où $\partial_{int}\Omega$ sont les frontières internes (fluide/solide), $\partial_g\Omega$ les frontières le long de la grille qui réalise le pavage du domaine et $\partial_{ext}\Omega$ les frontières externes de l'ensemble du domaine Ω .

Une méthode séquentielle par blocs en deux étapes, décrite dans la figure 6.2, est développée pour prédire l'écoulement interstitiel dans des coupes histologiques 2D d'ostéosarcomes. Plus d'étapes peuvent être introduites sans difficulté supplémentaire mais ne seront pas présentées ici. De même, cette étude a été menée pour caractériser les propriétés élastiques du milieu, mais ce point ne sera pas abordé ici.

Trois échelles sont utilisées et L_1 , L_2 et L_3 sont ordonnées de telle sorte que $L_1 < L_2 < L_3$. Comme aucune hypothèse n'est faite sur la séparation des échelles (pas de plateau sur le calcul des porosité par exemple, cf chapitre 4), les longueurs ne sont pas caractéristiques et peuvent être choisies arbitrairement. Deux changements d'échelles successifs sont effectués, de l'échelle L_1 à l'échelle L_2 puis de l'échelle L_2 à l'échelle L_3 . Pour illustration ici seul le premier changement d'échelle est décrit.

À l'échelle du pore L_1 , où les phases fluide et solide sont toutes deux considérées explicitement, l'écoulement de fluide interstitiel est caractérisé par des effets visqueux dominants correspondant à un écoulement avec un faible nombre de Reynolds. Le fluide est supposé être newtonien et incompressible. Par conséquent, l'écoulement entraîné par un gradient de pression ∇p_{L_1} , est modélisé par les équations sur Ω qui correspondent au problème de Stokes suivant :

$$\begin{cases} \mu \nabla^2 \mathbf{u}_{L_1} - \nabla p_{L_1} = -\rho \mathbf{g}, & \text{on } \Omega_f \text{ (} L_1\text{-scale)} \\ \nabla \cdot \mathbf{u}_{L_1} = 0, & \text{on } \Omega_f \\ \mathbf{u}_{L_1} = 0, & \text{on } \partial_{int}\Omega \\ \text{Boundary conditions (BCs) on } \partial_g\Omega \text{ and } \partial_{ext}\Omega \end{cases} \quad (6.1)$$

où μ et ρ sont la viscosité et la densité du fluide, respectivement, et u_{L_1} et g sont la vitesse du fluide et la gravité, respectivement. Dans ce qui suit, le terme de gravité n'est pas considéré car, d'une part, son rôle supposé négligeable sur l'écoulement interstitiel dans l'ostéosarcome, et d'autre part, le champ de pression hydrostatique résultant peut être inclus dans le terme de pression ∇p_{L_1} .

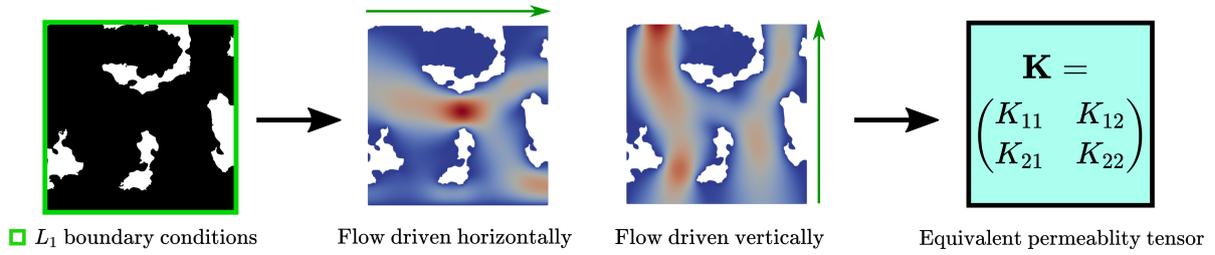


Figure 6.3: Image binaire d'un élément du pavage avec matrice osseuse en blanc, illustration de la magnitude de la vitesse du fluide interstitiel pour deux calculs représentatifs de l'écoulement de Stokes obtenus en imposant un gradient de pression horizontal et vertical de magnitude $\delta P = 1$, effectués pour calculer la perméabilité K à l'échelle L_2 sur un élément du pavage. (De Moreno et al. à soumettre).

Le premier changement d'échelle est basée sur la division du domaine initial en $N_x \times N_y$ sous-domaines, nommés A_i avec $i \in [[1, N_x \times N_y]]$, de même surface L_1^2 . Les tenseurs de perméabilité équivalente K_i à l'échelle L_2 sont calculés sur chaque sous-domaine A_i . Afin de construire un tenseur de perméabilité équivalent considérant l'anisotropie des champs sur un sous-domaine, il est nécessaire de considérer deux écoulements de Stokes entraînés par un gradient de pression de magnitude $(\delta P)^l$ dans chaque direction de l'espace, avec $l = 1$ pour l'horizontale et $l = 2$ pour la verticale (Figure 6.3). De plus, afin de déterminer explicitement les coefficients extra-diagonaux de tenseur équivalent, nous considérons les vitesses moyennes (6.2), fournies ici sur un sous-bloc nommé A pour des raisons pratiques.

$$\langle \mathbf{u}_{L_1} \rangle_A = \frac{1}{|A|} \int_{A_f} \mathbf{u}_{L_1} dS \quad (6.2)$$

où $|A|$ correspond à la surface d'un des éléments de la grille et A_f correspond au domaine fluide de A .

Ensuite le tenseur de perméabilité équivalent à l'échelle L_2 , pour un sous-domaine A est donné par l'équation (6.3) d'après [91]:

$$(K_{kl})_{1 \leq k, l \leq 2} = \left(-\mu \langle \mathbf{u}_{L_1, k} \rangle_A^l \frac{L_1}{(\delta P)^l} \right)_{1 \leq k, l \leq 2} \quad (6.3)$$

où l correspond à la direction de l'écoulement testé, et k correspond to la direction de l'espace.

À l'échelle L_2 , l'écoulement est maintenant décrit par un géométrie équivalente et constante par morceaux $\tilde{\Omega}$, par l'équation de Darcy incompressible (6.4) avec $\partial \tilde{\Omega}$ la limite. Le tenseur de perméabilité $\mathbf{K}(s)$ est constant par morceaux (s étant la variable d'espace à l'échelle L_2) avec des valeurs sous chaque sous-bloc correspondant à \mathbf{K}_i avec $i \in [[1, N_x \times N_y]]$. La vitesse \mathbf{U}_{L_2} est la vitesse de Darcy et P_{L_2} la pression à l'échelle L_2 .

$$\begin{cases} \mathbf{U}_{L_2} = -\frac{\mathbf{K}(s)}{\mu} \cdot \nabla P_{L_2}, & \text{on } \tilde{\Omega} \text{ (} L_2\text{-scale)} \\ \nabla \cdot \mathbf{U}_{L_2} = 0, & \text{on } \tilde{\Omega} \\ \text{BCs on } \partial \tilde{\Omega} \end{cases} \quad (6.4)$$

Ce résultat est motivé par celui de l'homogénéisation classique [96] sur une géométrie périodique.

Ce processus peut être reproduit par un nouveau changement d'échelles de Darcy vers Darcy. Dans le détails des travaux, différents types de conditions limites ont été testées telles que le non glissement la symétrie et l'imposition d'une condition aux limites de pression linéaire. N'étant pas au coeur de mon projet de thèse je ne décrirai pas cette partie de l'étude, cependant après une étude de la convergence en terme de conditions aux limites et de taille du domaine, il a été choisi d'utiliser des conditions aux limites symétriques aux bords des éléments du pavage et

$L_1 = 800px$, $L_2 = 1600px$ et $L_3 = 8000px$. Au final les résultats obtenus ont la forme suivante figure 6.4:

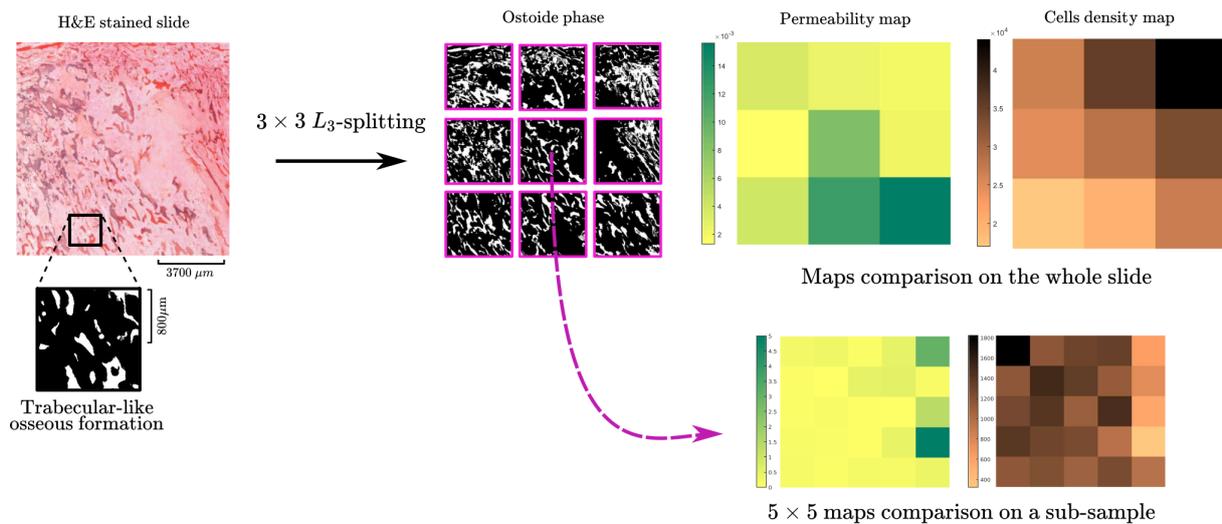


Figure 6.4: Comparaison des cartes de perméabilité (px^2) et densité cellulaires (mm^{-2}) sur une coupe histologique marquée H&E. (De Moreno et al. à soumettre)

A partir de ces cartes (figure 6.4), des calculs de corrélation peuvent être par exemple effectués reliant la densité cellulaire à la perméabilité. Ici la densité cellulaire est une mesure de réponse au traitement au sens où un faible nombre de cellules caractérise une zone de bonne réponse alors qu'une densité cellulaire importante correspond à une zone de mauvaise réponse. Dans le cadre de l'exemple illustré ici, les zones d'un calcul de corrélation de Spearman sur un découpage 15×15 (illustré par le découpage fin) donne un coefficient de -0.493 et une p -valeur de 0.013 . Ce résultat illustré sur les deux cartes du bas met en évidence que les zones de bonne réponse sont associées à une perméabilité plus importante. Dans cette étude menée sur une cohorte de 5 patients, nous avons pu montrer que le signe de la corrélation était dépendant de la micro-structure osseuse en structure trabéculaire ou en forme de dentelle. Cette dépendance à la micro-structure met certainement en évidence que les couplages mécano-biologiques dans ces différentes zones sont de natures différentes, un point qui pourra être mis en évidence dans l'avenir, entre autre, par l'exploration de coupe immunohistologiques et de l'environnement immunitaire des tumeurs. Dans ce travail, le même type d'approche a été développée pour caractériser les propriétés équivalentes élastiques d'un double milieu poreux (phase osseuse rigide et phase environnante très souple), un aspect qui n'a pas été développé dans cette thèse mais dont les résultats sont obtenus grâce au même type de traitement d'images présenté ici. Ces nouveaux quantificateurs mécano-biologiques pourraient dans l'avenir nous permettre d'obtenir de nouveaux mécano-marqueurs de la réponse aux traitements, spécifiques aux patients et pouvant être intégrés dans une stratégie d'adaptation thérapeutique.

Conclusion et perspectives

Dans cette thèse, des techniques de segmentation robustes d'images cliniques histologiques et immunohistologiques ont été développées pour s'efforcer de répondre aux questions ouvertes sur l'évolution des ostéosarcomes et leur réponse aux traitements. Ces techniques ont permis la détermination de métriques hétérogènes dans l'espace et portant sur la micro-architecture et la distribution des cellules dans le micro-environnement tumoral. Permettant pour la première fois de corréler spatialement la matrice cellulaire aux cellules immunitaires du micro-environnement, ces méthodologies sont également essentielles pour la mise en oeuvre d'une nouvelle approche d'imagerie fonctionnelle numérique obtenue à partir d'images de résolution microscopique et basées sur la théorie du transport et de l'élasticité en milieux poreux. Un logiciel numérique interne a été développé et validé pour la segmentation et le pavage des images.

D'un point de vue méthodologique, la littérature sur la caractérisation de la corrélation entre milieux continus et discrets ainsi que de l'exploration de propriétés mécaniques équivalentes, en milieux poreux hautement hétérogènes avec une résolution à l'échelle du pore, est rare. Cette thèse propose de nouveaux modèles et des développements numériques qui peuvent être adaptés à d'autres applications dans la recherche en milieux poreux. Les résultats obtenus sur une cohorte de 8 patients montrent les possibilités qu'offrent ces nouvelles techniques en terme de caractérisation des cancers osseux à génomique complexe, de suivi des patients et d'adaptation personnalisée des traitements.

Dans le premier chapitre, une revue non exhaustive des connaissances actuelles sur l'ostéosarcome est effectuée à plusieurs échelles spatiales. Une attention particulière est accordée à la description du micro-environnement de l'ostéosarcome, une échelle à laquelle cette tumeur osseuse primaire peut être vue comme un milieu poreux hétérogène. L'accent est porté sur l'hétérogénéité spatiale de l'ostéosarcome et les réponses cliniques que peut apporter son exploration. Les objectifs de la thèse sont ensuite introduits.

Le second chapitre propose une stratégie de segmentation d'images cliniques, issues de coupes histologiques et immunohistologiques, par des techniques de machine learning. Les techniques proposées s'appuient sur une information basée sur la couleur des images.

Le troisième chapitre a pour objectif d'améliorer cette segmentation par l'utilisation de filtres qui explorent la texture des images. S'appuyant sur la difficulté d'interpréter les filtres issus de techniques de machine learning, du type Support Vector Machine, des développements théoriques ont été effectués pour projeter ces filtres efficaces mais non interprétables sur des bases de filtres linéaires aux dérivées partielles. Les cas d'images idéalisées ou d'image bruitées sont explorés. De plus, une exploration des filtres invariants par rotation ou symétrie axiale a permis de mettre en évidence de nouveaux filtres linéaires et non linéaires pertinents pour l'étude des ostéosarcomes.

Le quatrième chapitre concerne le développement de métriques permettant la caractérisation de l'hétérogénéité spatiale du micro-environnement tumoral. Un intérêt particulier est porté à l'étude des corrélations entre phase discrète associée aux phases cellulaires et continue associée à la matrice extra-cellulaire. Une approche innovante de calcul de corrélation locales est proposée.

Le chapitre 5 est dans la continuité de ces développements et s'intéresse aux corrélations

locales entre porosité et densité cellulaire immunitaire. Les résultats retrouvent et quantifient certaines informations de la littérature comme l'aspect immunosuppresseur de l'ostéosarcome, en particulier dans les zones de mauvaise réponse au traitement mais également apporte de nouvelles conclusions comme la prévalence de la présence de macrophages CD163 dits "anti-tumoraux" dans les régions osseuses de bonnes réponses aux traitements. Ces conclusions ouvrent en particulier de nouvelles perspectives quand à l'exploration de nouvelles cibles thérapeutiques pour l'ostéosarcome.

Le sixième chapitre résume la manière dont les développements de cette thèse ont permis de mettre en oeuvre de nouvelles techniques d'imageries fonctionnelles numériques à partir de coupes histologiques. Cette approche basée sur la théorie du transport et de l'élasticité en milieux poreux ouvre une nouvelle voie quant à l'exploration de mécano-marqueurs pour la caractérisation du rôle des effets mécaniques dans le développement de l'ostéosarcome, le suivi des patients et l'adaptation des traitements.

En raison de la rareté de l'ostéosarcome qui représente 350 cas en France par an, et du nombre restreint de mauvais réponders avec une réponse mixte spatialement hétérogène (25% de patients sur la cohorte de Toulouse), cette étude a été menée sur une cohorte de taille limitée comme preuve de concept. L'extension de ce travail est d'élargir la cohorte pour confirmer les résultats de la corrélation entre la réponse au traitement, la distribution des cellules immunitaire et la microstructure de la matrice extra-cellulaire osseuse.

Par ailleurs, comme illustré dans le chapitre 5, les coupes immunohistologiques présentent de nombreux biais liés à la difficulté de manipuler des tissus minéralisés. On peut observer entre autre, des déchirures de tissus, des variations dans l'intensité des marquages ou encore des défauts de marquages. Des études sont en cours au CRB-Cancer Toulouse pour améliorer la qualité des coupes, en particulier dans le processus de déminéralisation des tissus qui conditionne l'intensité des marquages mais également l'intégrité du tissu lors de son dépôt sur la lame de verre.

Bien que l'approche développée introduise de nombreux concepts de machine learning et de milieux poreux, les résultats, c'est-à-dire cartes de porosité, de densité cellulaire, cartes de corrélations locales et coefficients de corrélation, sont simples à manipuler. Nous espérons qu'à court terme, ces développements numériques pourront être utilisés par les biologistes et les cliniciens pour étudier l'ostéosarcome et explorer de nouvelles pistes de l'amélioration des diagnostics et des stratégies thérapeutiques. En particulier, la méthodologie d'imagerie fonctionnelle numérique peut être directement appliquée aux images immunohistologiques, et les changements de polarisation des macrophages (changement de fonction de pro à anti-tumoral) dus aux effets mécaniques pourraient être explorés à l'échelle du tissu. Ceci est possible grâce aux techniques de segmentation développées et dédiées à ces images. Cette approche pourrait donc aider à trouver de nouvelles cibles pour les traitements d'immunothérapie des ostéosarcomes.

L'optimisation numérique est en cours pour atteindre des temps de calcul plus courts, diminuer l'intervention de l'expert et augmenter le potentiel d'utilisation du logiciel développé par des réorganisations modulaires et ergonomiques du logiciel.

A plus long terme, la prise en compte du réseau vasculaire pourrait être mise en oeuvre. Ceci complexifie le problème en introduisant un réseau dans un milieu poreux, cependant la prise en compte de la vascularisation est en particulier essentielle pour étudier le rôle de l'hypoxie dans le développement de l'ostéosarcome et la réponse immunitaire associée.

De plus, l'acquisition de données tridimensionnelles à l'échelle de la cellule doit être envisagée pour résoudre le problème de manière plus complète. Des techniques de recalage permettent à l'heure actuelle de reconstruire un milieu 3D à partir de coupes histologiques et immunohistologiques [97], cependant ces techniques d'imagerie détériorent le milieu et des acquisitions 3D s'appuyant sur des développements récents (tomographie à rayons X haute résolution) pour-

raient permettre de visualiser le milieu osseux comme la phase cellulaire en 3D. Pour ce faire, des marqueurs de cellules pertinents doivent être identifiés pour cette technologie, car l'acquisition de telles données dans les tissus minéralisés reste un défi. Le logiciel d'imagerie fonctionnelle est lui aussi adaptable aux données 3D et l'approche séquentielle pourra être très puissante pour réduire le temps de calcul.

En conclusion, les développements de cette thèse visent à intégrer des paramètres géométriques et biomécaniques objectifs et quantifiables pour l'exploration de tissus cancéreux hétérogènes dans un but de stratification des patients et d'une meilleure prise en charge thérapeutique. Ce travail se situe au croisement entre des sujets de grand intérêt pour la recherche sur le cancer tels que l'étude l'interaction cellule/MEC, la quantification de l'hétérogénéité dans les tumeurs, le rôle de la mécanique dans les tumeurs. La méthodologie est générique et peut être adaptée à d'autres types d'images, de tumeurs et de problèmes physiques.

Appendix A

Annexe 1: Recalage spatial

A.1 Recalage spatial

Pour une même pièce de résection d'un patient, plusieurs coupes histologiques et immuno-histologiques peuvent être effectuées.

Une première raison est d'observer différentes populations cellulaires par différents marquages. Dans l'immuno-histologie simplexes, on ne peut pas appliquer différents marquages sur une même lame. Les coupes successives de tissu sont déposées à la main sur les lames en verre, ce qui entraîne des orientations différentes des coupes histologiques sur les images scannées : Deux même coordonnées de pixels ne correspondent pas à la même localisation sur deux lames de tissus. Le recalage spatial entre coupes histologique est donc un problème récurrent [97]. Certaines nouvelles techniques de marquages, telles que le multiplex, permettent de marquer différentes populations cellulaires sur une même lame mais à l'heure actuelle ne permettent pas d'identifier la phase osseuse. De plus, ces techniques restent très coûteuses et ne sont pour l'instant pas utilisées en routines dans les hopitaux.

Une autre raison a été la tentative de faire de la reconstruction 3D du profil de matrice osseuse et du réseau vasculaire par coupes immuno-histologiques avec marquage au CD34 (endothélium vasculaire).

A.1.1 Principe du recalage spatial

Soient trois lames successives avec leur marquage respectif A, B et C . On choisit une des trois images comme image de référence, par exemple A .

Le recalage spatial entre A et B est une application φ_B qui va associer à chaque coordonnées de pixels de l'image B une estimation de la coordonnées du pixel correspondant dans l'image A , $i_A, j_A = \varphi_B(i_B, j_B)$ (figure A.1).

Il n'est pas question ici de recréer une image B déformé tel qu'elle s'ajuste à A , car cela prendrait inutilement trop de place en mémoire.

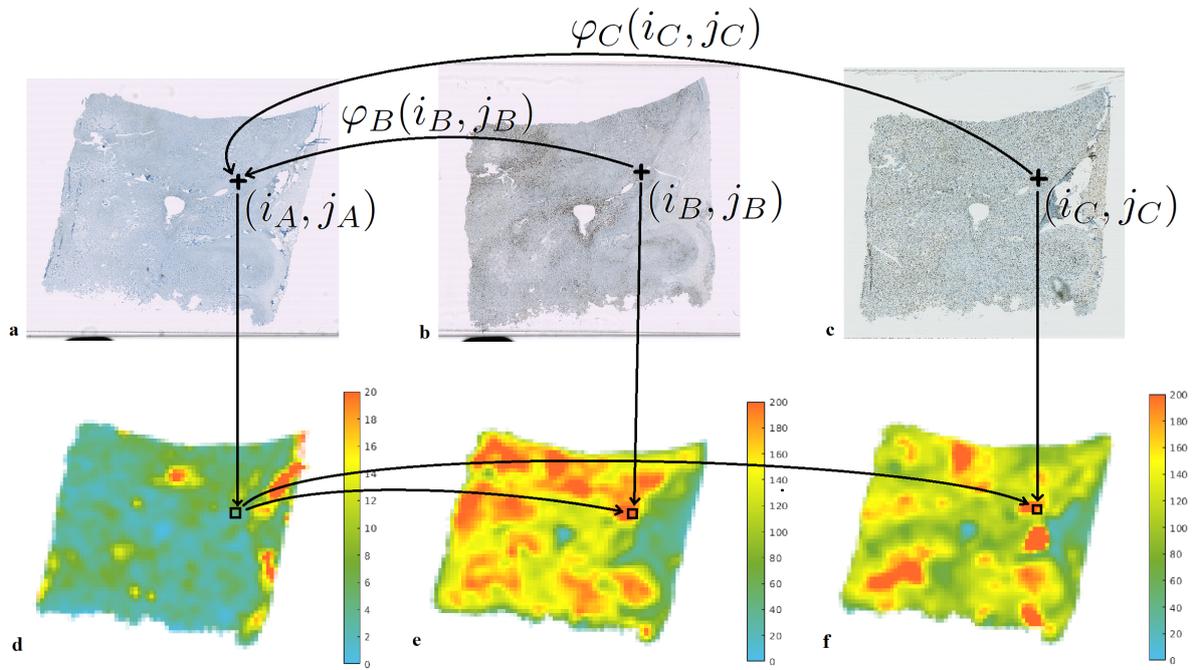


Figure A.1: Schema illustratif du principe de recalage entre trois lames successives et cartes de densité cellulaire recalées sur le CD8, a) et d) CD8, b) et e) CD68 et c) et f) CD163

Si les images B et C sont recalées sur A , alors il est aussi possible de calculer des corrélations entre B et C .

Lors du calcul des cartes sur l'image B , le pixel de coordonnées (i_B, j_B) n'intervient pas dans la fenêtre déterminée par (i_B, j_B) mais dans celle associée à $\varphi_B(i_B, j_B)$. Cette fenêtre pour B correspond à la même "zone géographique" de l'image A , du point de vue du tissu biologique. Ainsi, pour les mêmes coordonnées de fenêtres entre les cartes de A et B , les quantités peuvent être comparées car elles sont calculées sur des zones quasiment identiques du positionnement qu'avait les tissus dans la tumeur.

La figure A.2 montre un exemple, sur des fenêtres volontairement grossières, la correspondance entre les zones d'une lame de tissu à l'autre.

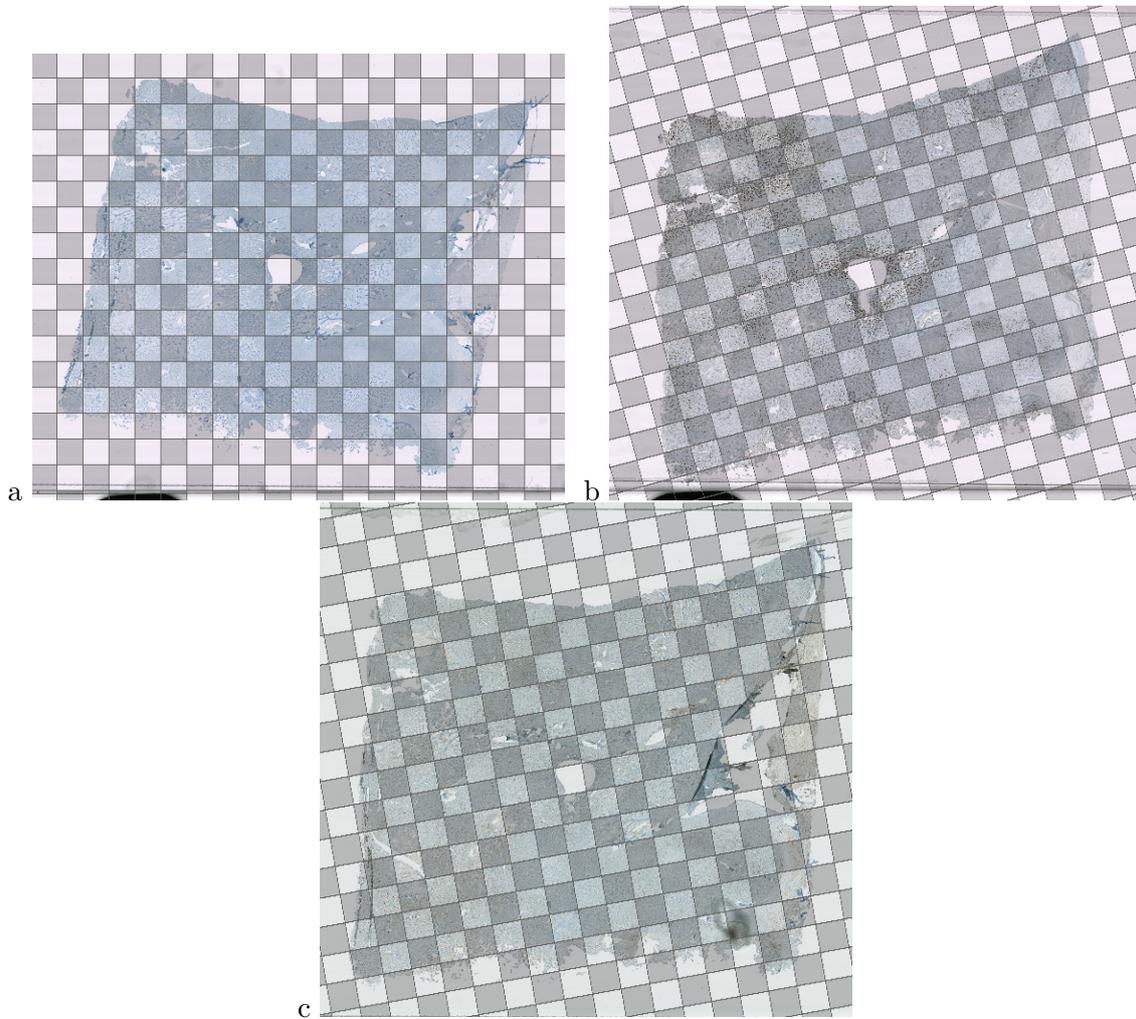


Figure A.2: Illustration du recalage spatial sur trois coupes immuno-histologiques avec CD8 comme marquage de référence, a) CD8, b) CD68 et c) CD163

Des méthodes de Cross-correlations sur des éléments précis à petite échelle sont communément utilisées dans les cas de suivi temporel ou de déformation de tissu, par exemple pour en étudier l'élasticité. Mais dans notre cas, d'une lame à l'autre, les cellules ne sont pas les mêmes et donc les noyaux visibles ne sont pas associables deux à deux. C'est pourquoi nous devons nous appuyer sur des méthodes et des quantités plus globales.

A.1.2 Recalage spatial basé sur le masque du tissu

Lorsque le bord du tissu de chaque lame des coupes immuno-histologiques successives ne sont pas trop détériorés (plis, partie arrachées etc), il est possible d'appliquer un recalage spatial affine s'appuyant sur le masque. L'avantage de se recalage spatial est qu'il est automatisé, il ne demande pas l'aide de l'utilisateur.

La figure A.3 montre le cas du patient 14R004225 pour lequel les bords des tissus sont suffisamment propres et similaires.

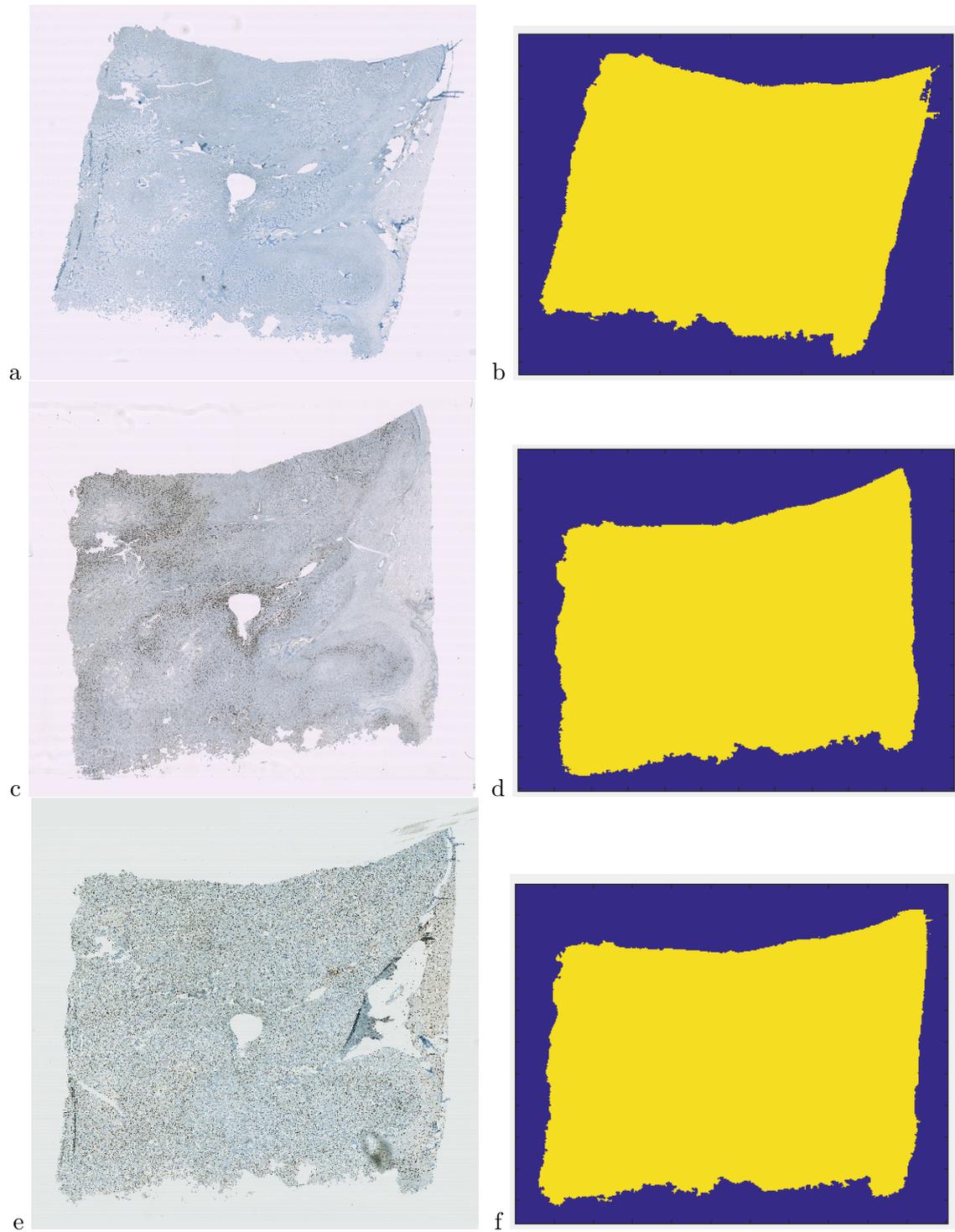


Figure A.3: Coupes immuno-histologiques et masques associés, a) et b) CD8, c) et d) CD68, e) et f) CD163

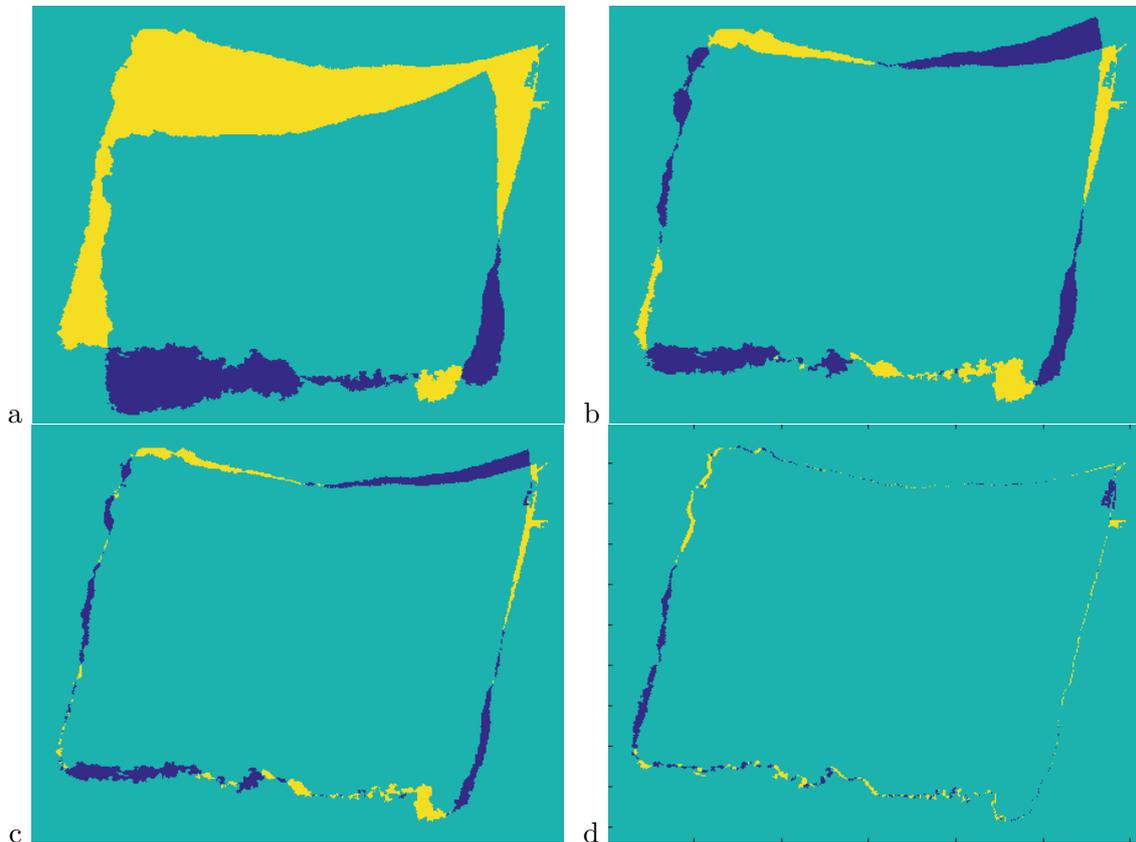


Figure A.4: Recalage spatial basé sur le masque entre deux lames successives marquées, CD68 recalé sur le CD8

La méthode utilisée est une variante de descente de gradient dont la fonction à minimiser est une erreur quadratique sur les fonctions caractéristiques des masques. Un produit de convolution avec une gaussienne peut-être appliqué sur les fonctions caractéristiques des masques, de manière à éliminer le bruit dû aux quelques différences de pixels entre les masques.

A.1.3 Recalage spatial avec points de repère

Lorsque les bords des lames de tissus ne permettent pas un recalage basé sur les masques, nous pouvons utiliser une deuxième méthode basée sur des points de repère déterminés par l'utilisateur. Cette méthode, contrairement à la précédente, demande une intervention de l'utilisateur.

On cherche ici à trouver une régression 2D d'une lame associée, à un marquage, à une lame de référence, c'est-à-dire une application φ tel que l'erreur quadratique soit minimisée sur des points de repère.

Lors des recalages par masque qui sont affines, nous avons pu constater que malgré une très bonne approximation d'une transformation affine pour le recalage entre lame, il y avait une légère déformation supplémentaire, probablement dû à l'élasticité des coupes de tissu. C'est pourquoi, nous avons augmenté le degré pour rechercher des transformations quadratiques.

Soient une suite de n points $(X_k)_k$, avec X_k de coordonnées $(X_{1,k}, X_{2,k})$, sur l'image B que l'on veut faire correspondre à la suite de points $(Y_k)_k$ sur l'image de référence A .

$$\varphi(X_k) = b + A.X_k + Q : (X_k \otimes X_k) \quad (\text{A.1})$$

où b est un vecteur 2×1 , A est un vecteur 2×2 , Q est un tenseur $2 \times 2 \times 2$ avec $Q_{(1,1,2)} = Q_{(1,2,1)}$ et $Q_{(2,1,2)} = Q_{(2,2,1)}$.

On cherche à minimiser l'erreur :

$$\left\| \sum_{k=1}^n \varphi(X_k) - Y_k \right\|^2 \quad (\text{A.2})$$

Il y a alors 12 inconnus à déterminer qu'on peut ranger dans deux vecteurs colonnes :

$$C_1 = \begin{pmatrix} b_1 \\ A_{1,1} \\ A_{1,2} \\ Q_{1,1,1} \\ Q_{1,1,2} \\ Q_{1,2,2} \end{pmatrix} \quad C_2 = \begin{pmatrix} b_2 \\ A_{2,1} \\ A_{2,2} \\ Q_{2,1,1} \\ Q_{2,1,2} \\ Q_{2,2,2} \end{pmatrix} \quad (\text{A.3})$$

Soit la matrice M de taille $6 \times n$:

$$M = \begin{pmatrix} 1 & \dots & 1 \\ X_{1,1} & \dots & X_{1,n} \\ X_{2,1} & \dots & X_{2,n} \\ X_{1,1}^2 & \dots & X_{1,n}^2 \\ 2X_{1,1}X_{2,1} & \dots & 2X_{1,n}X_{2,n} \\ X_{2,1}^2 & \dots & X_{2,n}^2 \end{pmatrix} \quad (\text{A.4})$$

Nous avons alors les solutions :

$$C_1 = (M.M^T)^{-1}.M. \begin{pmatrix} Y_{1,1} \\ \vdots \\ Y_{1,n} \end{pmatrix} \quad \text{et} \quad C_2 = (M.M^T)^{-1}.M. \begin{pmatrix} Y_{2,1} \\ \vdots \\ Y_{2,n} \end{pmatrix} \quad (\text{A.5})$$

S'il n'y a pas au moins 6 points de repère, la matrice $M.M^T$ ne peut pas être inversible.

la figure A.5 montre un exemple de deux lames dont le bord ne permet pas un recalage basé sur le masque. On peut y voir un ensemble de 8 points sur chaque lames.

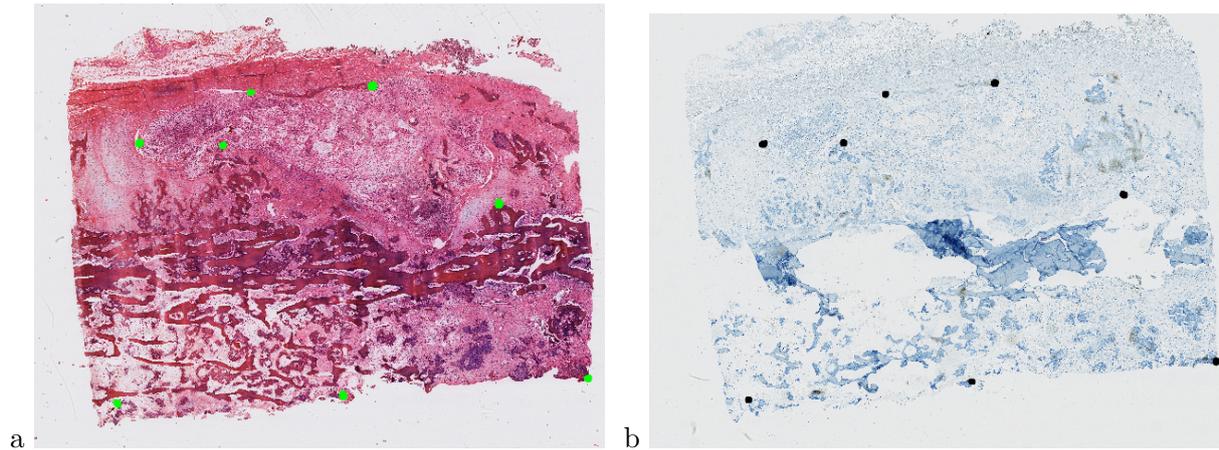


Figure A.5: Illustration de points de repère pour un recalage spatial basé sur des points de repère entre deux lames successives marquées, CD8 recalé sur le HE. a) HE avec points en vert, b) CD8 avec points en noir.

A.1.4 Exemple du recalage par points avec étude d'un cas sur des zones de lames

Les images du patient 19T014061 montrent une réponse très hétérogène au traitement (les lames de H&E et CD8 sont dans le figure A.5). Les lames de tissu étant également parfois détériorées (déchirure présente sur la lame de CD8, figure A.5 b)), nous avons décidé d'appliquer le recalage

par points et d'étudier les quantités sur des régions d'intérêt. Ces régions correspondent à des zones de bonne ou de mauvaise réponse, ainsi qu'une zone de réponse mixte.

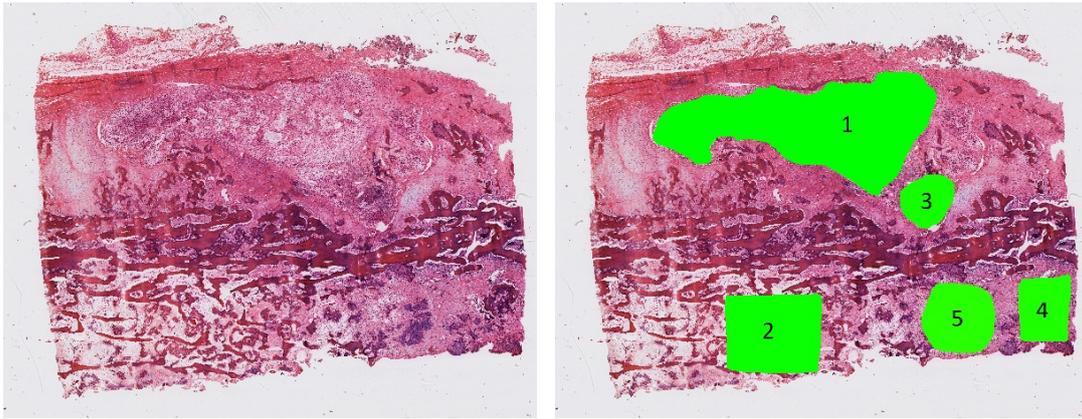


Figure A.6: Détermination manuelle de régions d'intérêt, 1,4,5 : mauvaise réponse au traitement, 2 : bonne réponse, 3 : réponse mixte region.

regions	1	2	3	4	5
H&E					
CD8					

Table A.1: Zoom sur les régions d'intérêt

regions	1	2	3	4	5
Porosity	0,97	0,8	0,76	0,88	0,89
Width of trabeculae (μm)	28	90	142	55	57
Density of stained cells (mm^{-2})	4,4	35,2	13,8	26,3	15,2
Density of not stained cells (mm^{-2})	1449	150	240	1456	1225
Percentage of stained cells	0,3%	19,0%	5,4%	1,8%	1,2%

Table A.2: Quantités sur les régions d'intérêt, coupe marquée au CD8.

Appendix B

Annexe 2: Coefficient de Spearman local

Dans ce qui précède les données ont été traitées par une approche de type Pearson. Le coefficient de corrélation de Spearman de variables X et Y est un coefficient de Pearson appliqué aux rangs des valeurs de X et Y ($rg(X)$ et $rg(Y)$) utilisé généralement lorsque les répartition des champs X relativement à Y ne sont pas linéaires.

$$r_s = \frac{Cov(rg(X), rg(Y))}{\sqrt{Var(rg(X))}\sqrt{Var(rg(Y))}} \quad (\text{B.1})$$

Comme précédemment, nous allons chercher à calculer ce coefficient localement dans l'espace.

Une première possibilité est de pondérer les valeurs en fonction de la distance, c'est-à-dire appliquer le Pearson local sur $rg(X)$ et $rg(Y)$.

Le problème de cette méthode est que :

Soit A une partie des données, et α_A la fonction caractéristique de A . Le résultat obtenu entre le coefficient de Spearman classique appliqué sur A et le coefficient de Spearman local avec comme poids le filtre α_A ne donne pas le même résultat. La raison est que le rang d'une valeur de X n'est pas le même selon qu'elle soit prise dans la partie A ou parmi toutes les valeurs de X .

Pour corriger ce problème et proposer un coefficient de corrélation de Spearman local plus cohérent, nous pouvons modifier la définition du rang.

Soit $X = [X_1, X_2, \dots, X_n]$ des données quelconques et $[X_{(1)}, X_{(2)}, \dots, X_{(n)}]$ les valeurs de X triées par ordre croissant.

La définition du rang habituel serait :

$$rg(X_{(1)}) = 1, \quad rg(X_{(k+1)}) = rg(X_{(k)}) + 1 \quad (\text{B.2})$$

Soit $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]$ les poids associés à X , et $\alpha_{(k)}$ le poids associé à $X_{(k)}$. On peut définir un rang pondéré par :

$$rg(X_{(1)}) = \alpha_{(1)}, \quad rg(X_{(k+1)}) = rg(X_{(k)}) + \alpha_{(k+1)} \quad (\text{B.3})$$

Appendix C

Annexe 3: Résultats complémentaires sur la cohorte de patients

Dans cette annexe, les coupes immunohistologiques (CD8, CD163, CD8), les cartes de porosité, les de densité cellulaire et les cartes de corrélation locales (porosité cellules marquées) sont données pour l'ensemble des patients disponibles dans la cohorte toulousaine ainsi que les coefficients de corrélation globale associés. Les patients dont les coupes étaient endommagées par des déchirures ou des marquages insuffisants ont été exclus.

C.1 Patients mauvais répondeurs

Patient 1

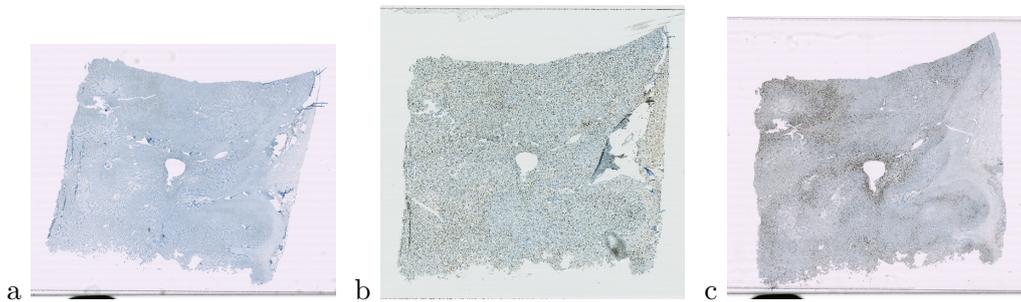


Figure C.1: Coupes immuno-histologiques a) marquage CD8, b) CD163, c) CD68

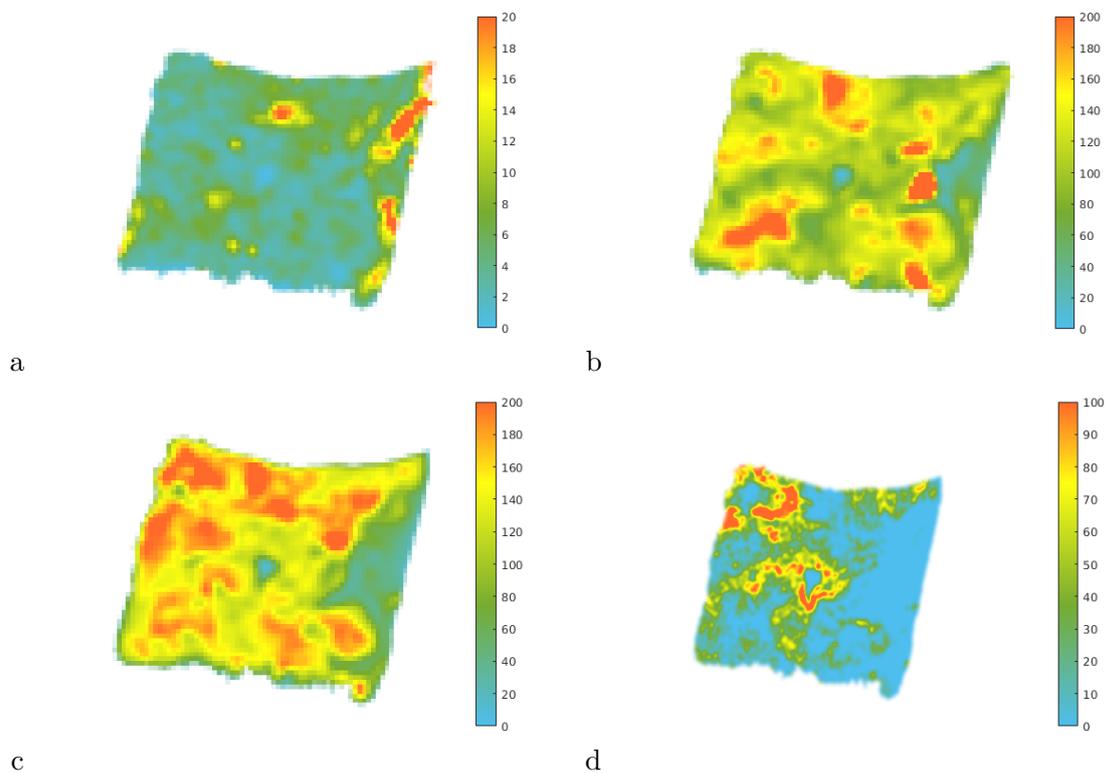


Figure C.2: Carte de densité cellulaire a) CD8, b) CD163, c) cellules totales CD68, d) grandes cellules CD68, cartes avec recalage spatial sur le CD8

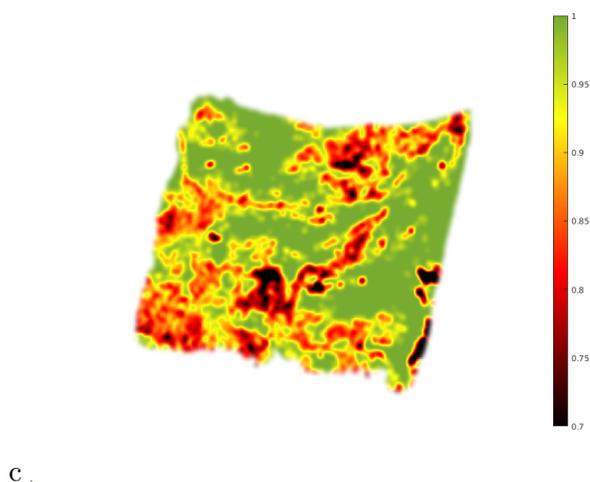


Figure C.3: Cartes de porosité, a) CD8, b) C163, c) CD68.

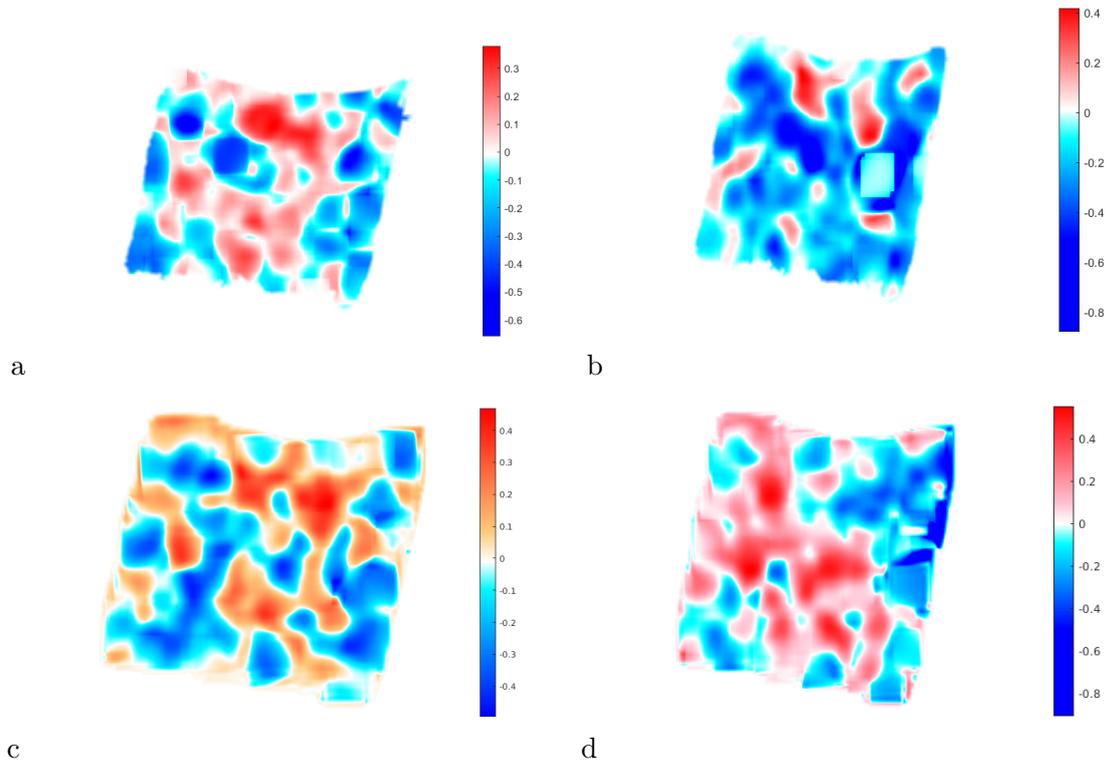


Figure C.4: Carte de corrélations locales porosité/densité cellulaire a) CD8, b) CD163, c) cellules totales CD68, d) grandes cellules CD68.

Patient 3



Figure C.5: Coupes immuno-histologique du patient 19T01 mauvais répondeur, a) marquage CD8, b) CD163, c) CD68.

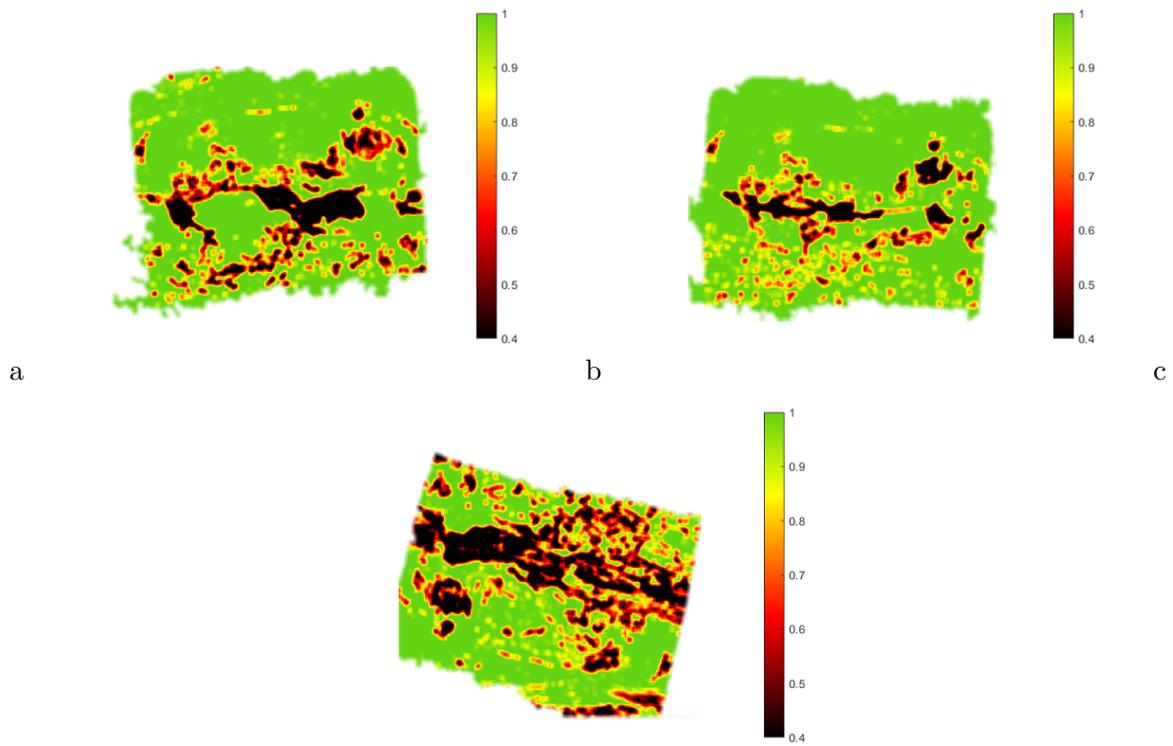


Figure C.6: Cartes de porosité a) CD8, b) CD163, c) CD68.

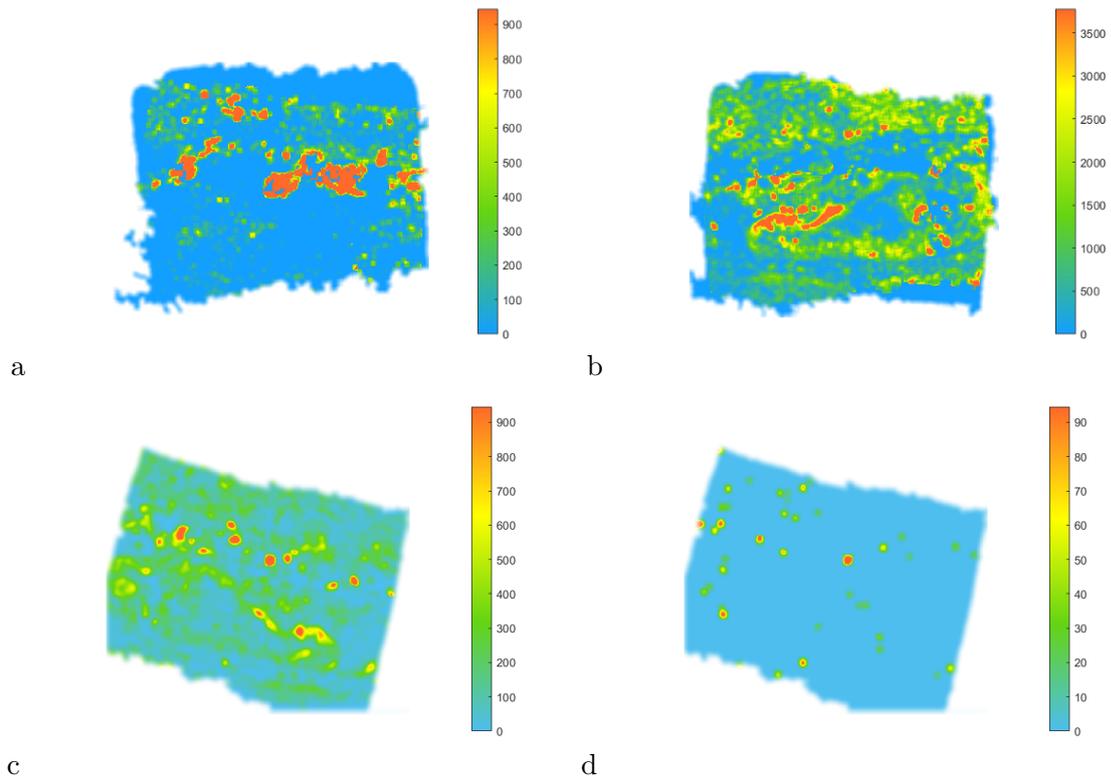


Figure C.7: Cartes de densités a) CD8, b) CD163, c) CD68 cellules totales, d) CD68 grandes cellules.

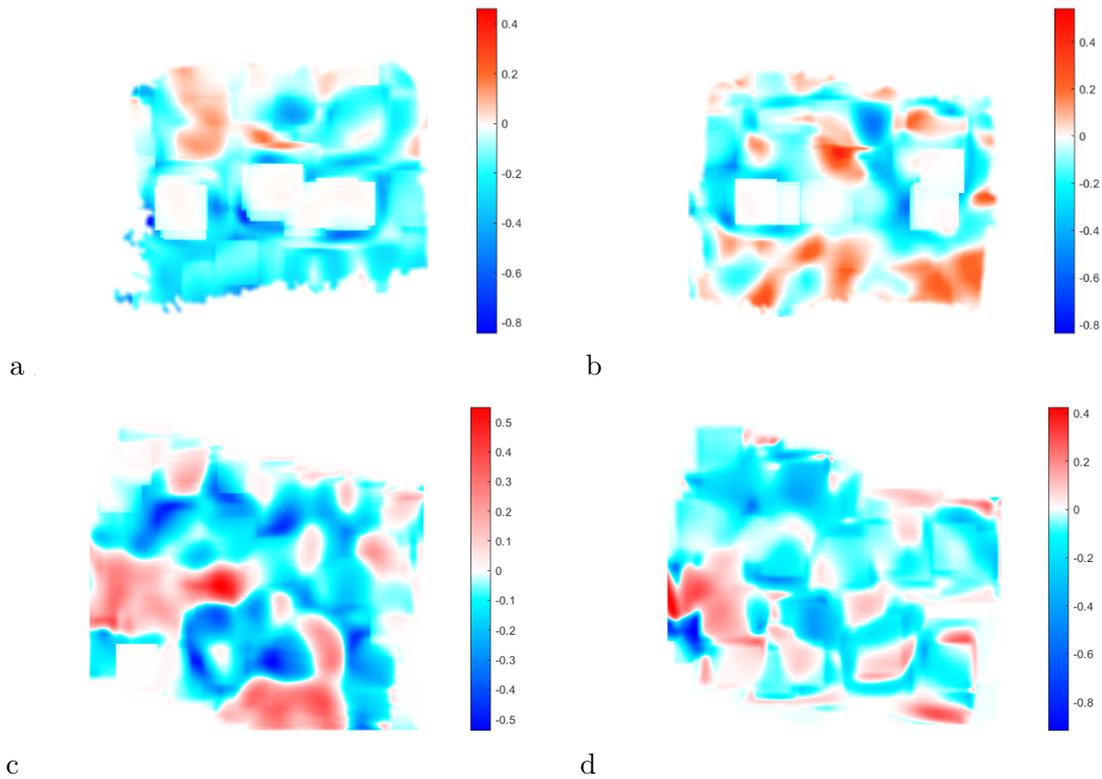


Figure C.8: Cartes de corrélations locales a) CD8, b) CD163, c) CD68 cellules totales, d) CD68 grandes cellules.

corrélation	CD8	CD163	CD68 t	CD68 g
Porosité	-0.462	-0.073	-0.117	-0.078

Table C.1: Corrélacion entre porosité et densité cellulaire, pour le CD68 : t = toutes cellules marquées, g = grandes cellules marquées

Dans cette section, 3 pièces de resections issues de patients mauvais répondeurs, selon le critère de Huvos Rosen, sont étudiées. Pour chacun des patients, les coupes immunohistologiques sont illustrées pour des marquages CD8 caractérisant les lymphocytes T, CD163 caractérisant des macrophages dont la présence est de bon pronostic pour les patients [16, 11], et le CD68 caractérisant les macrophages plutôt délétères au pronostic et incluant la mise en évidence des ostéoclastes [11].

C.2 Patients bons répondeurs

C.2.1 Patient 4



Figure C.9: Coupe immuno-histologique du patient 17T, a) marquage CD8, b) CD163, c) CD68

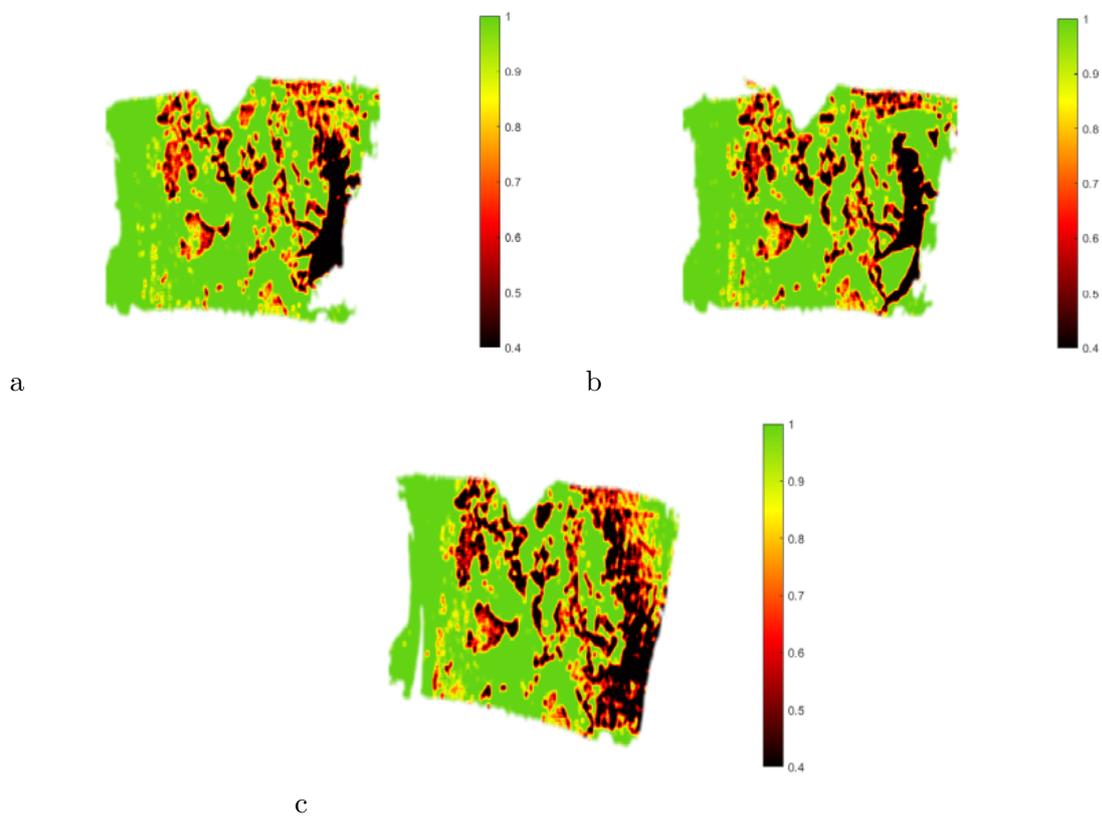


Figure C.10: Carte de porosité a) CD8, b) CD163, c) CD68.

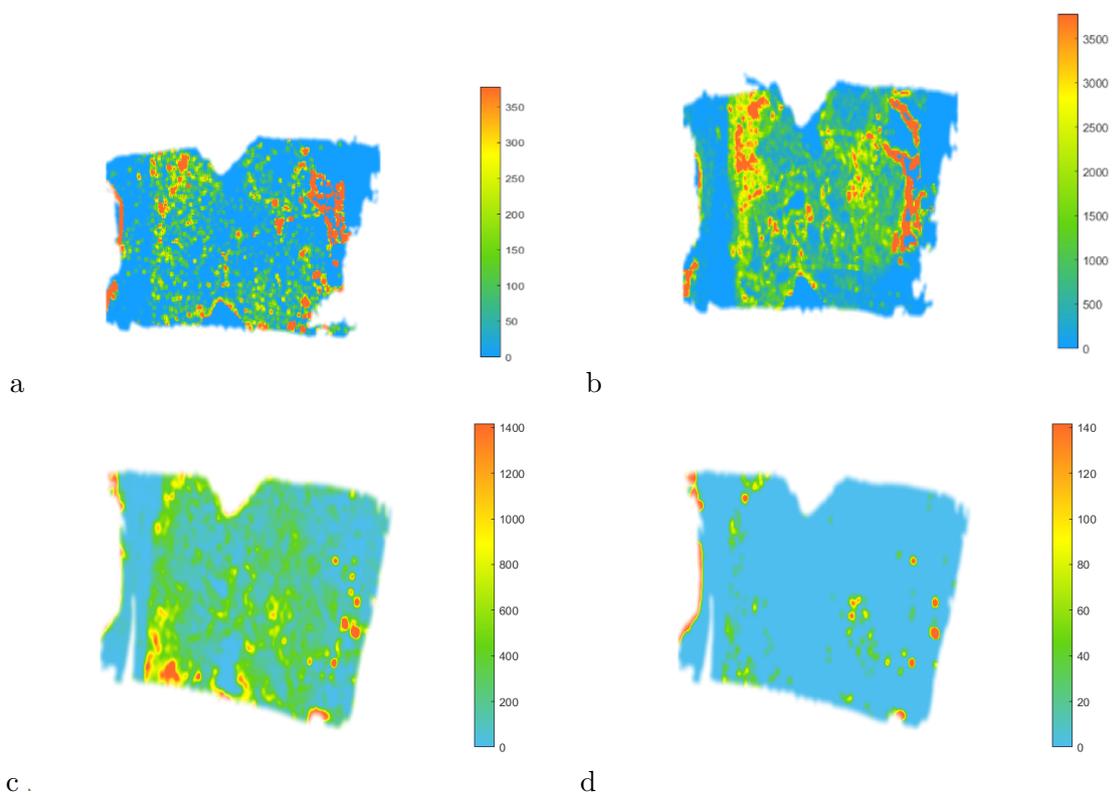


Figure C.11: Cartes de densité cellulaire a) CD8, a) CD163, c) CD68 cellules totales, d) CD68 grandes cellules.

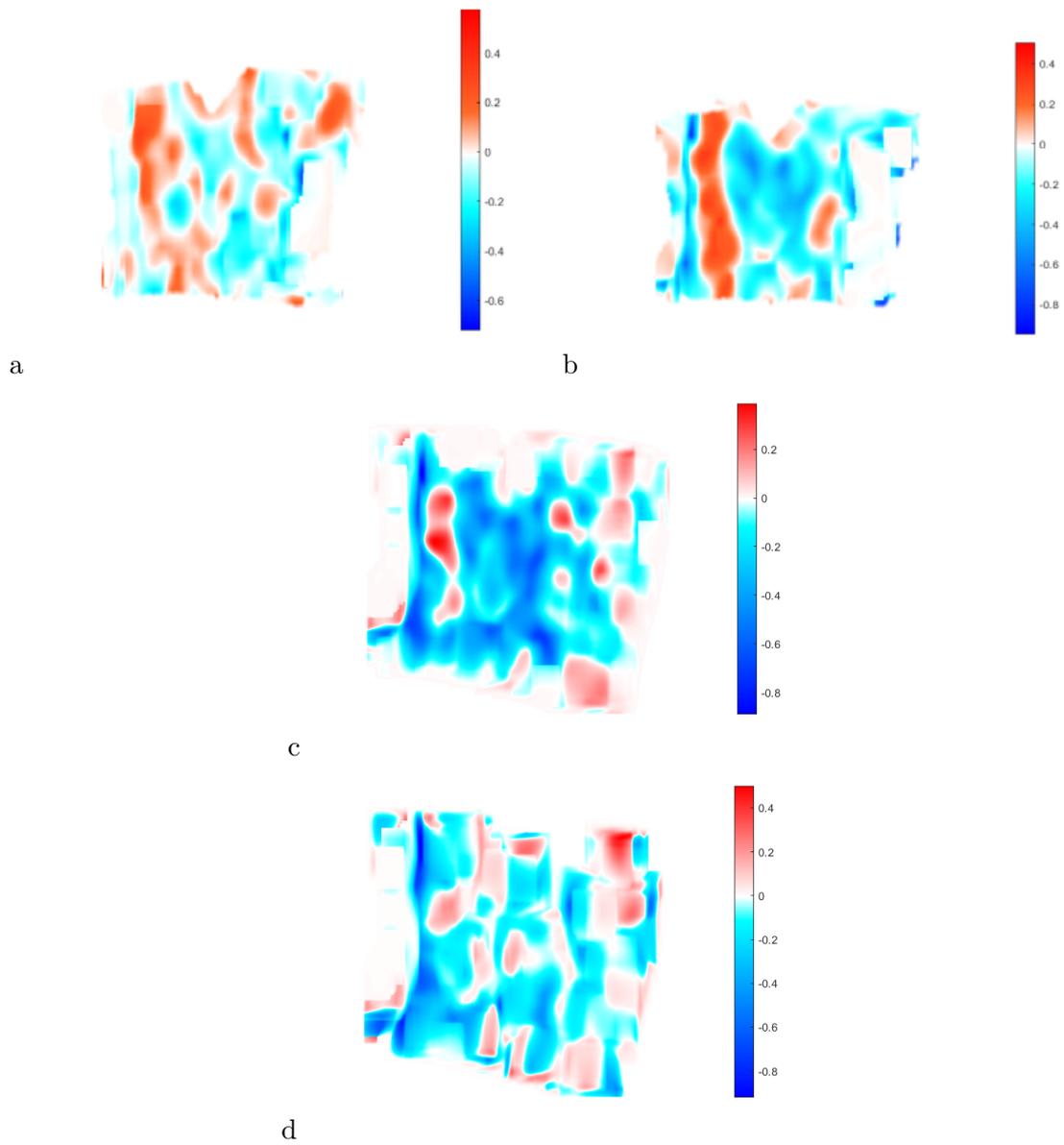


Figure C.12: Carte de corrélations locales a) CD8, b) CD163, c) CD68 cellules totales, d) CD68 grandes cellules.

corrélation	CD8	CD163	CD68 t	CD68 g
Porosité	-0.108	-0.228	-0.092	-0.033

Table C.2: Corrélacion entre porosité et densité cellulaire, pour le CD68 : t = toutes cellules marquées, g = grandes cellules marquées

C.2.2 Patient 5

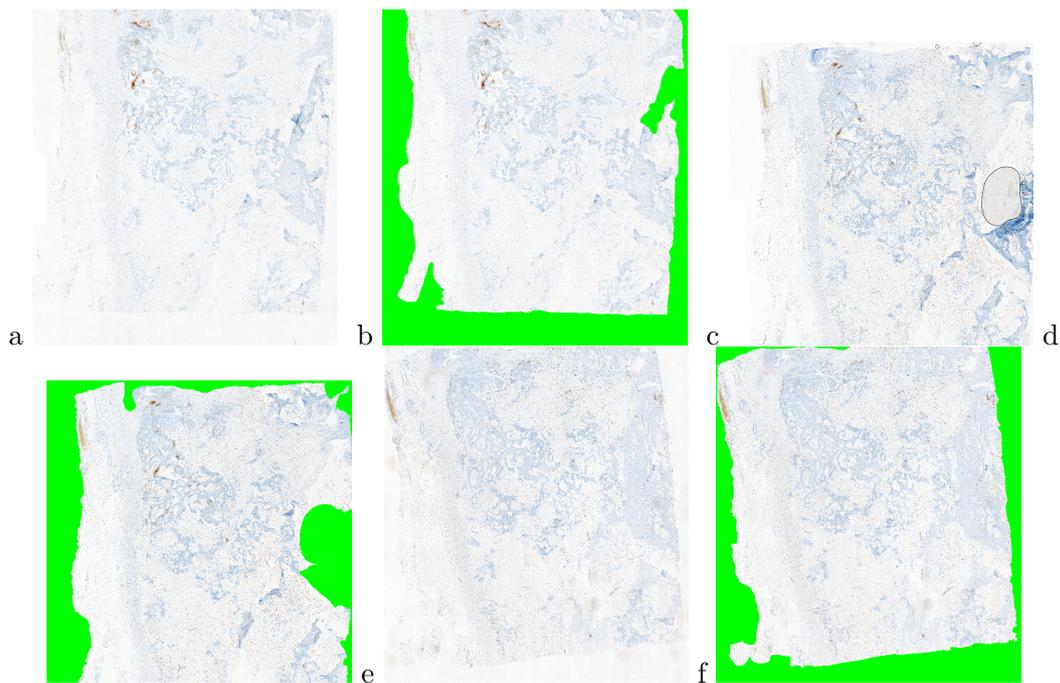


Figure C.13: Coupe immuno-histologique du patient 19T04, a) marquage CD8, b) CD8 masque, c) CD163, d) CD163 masque, e) CD68, f) CD68 masque.

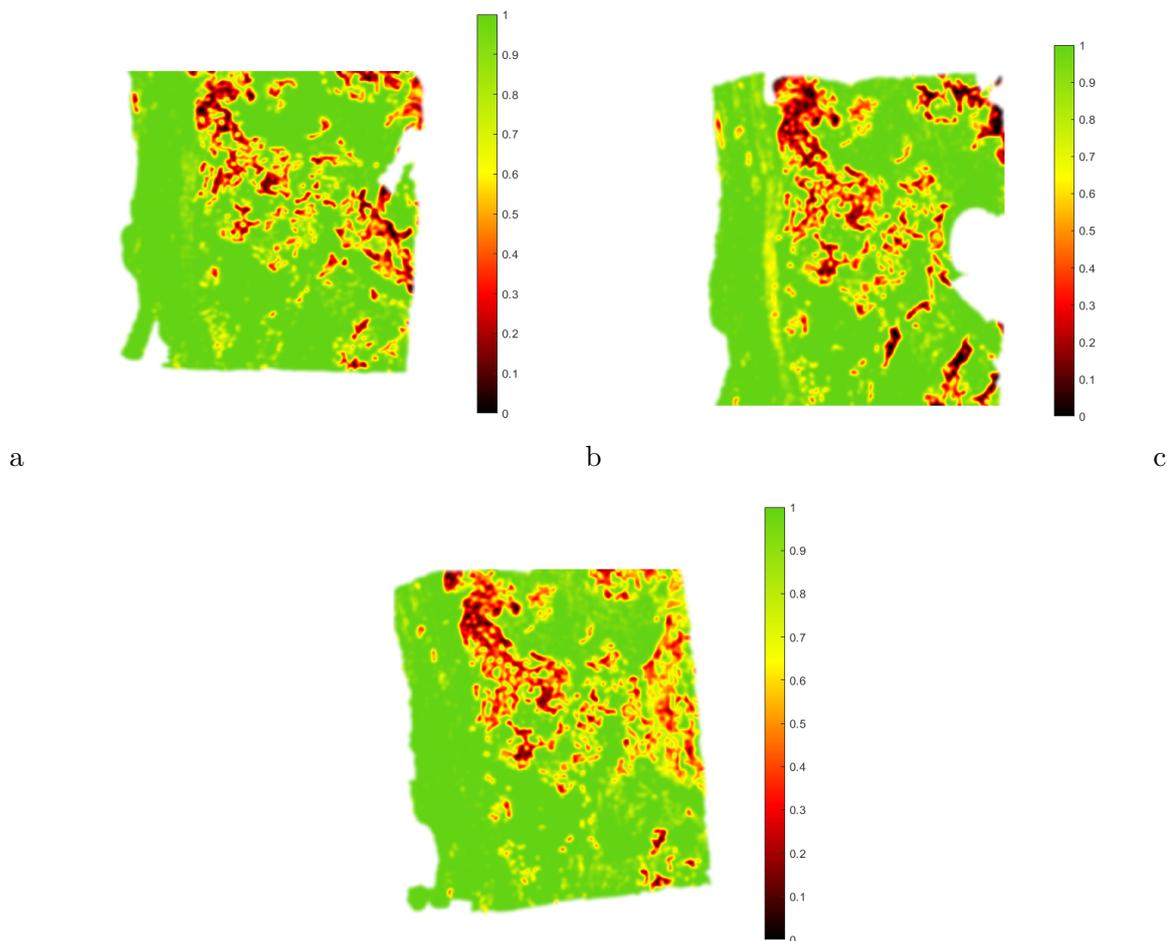


Figure C.14: Carte de porosité a) CD8, b) CD163, c) CD68.

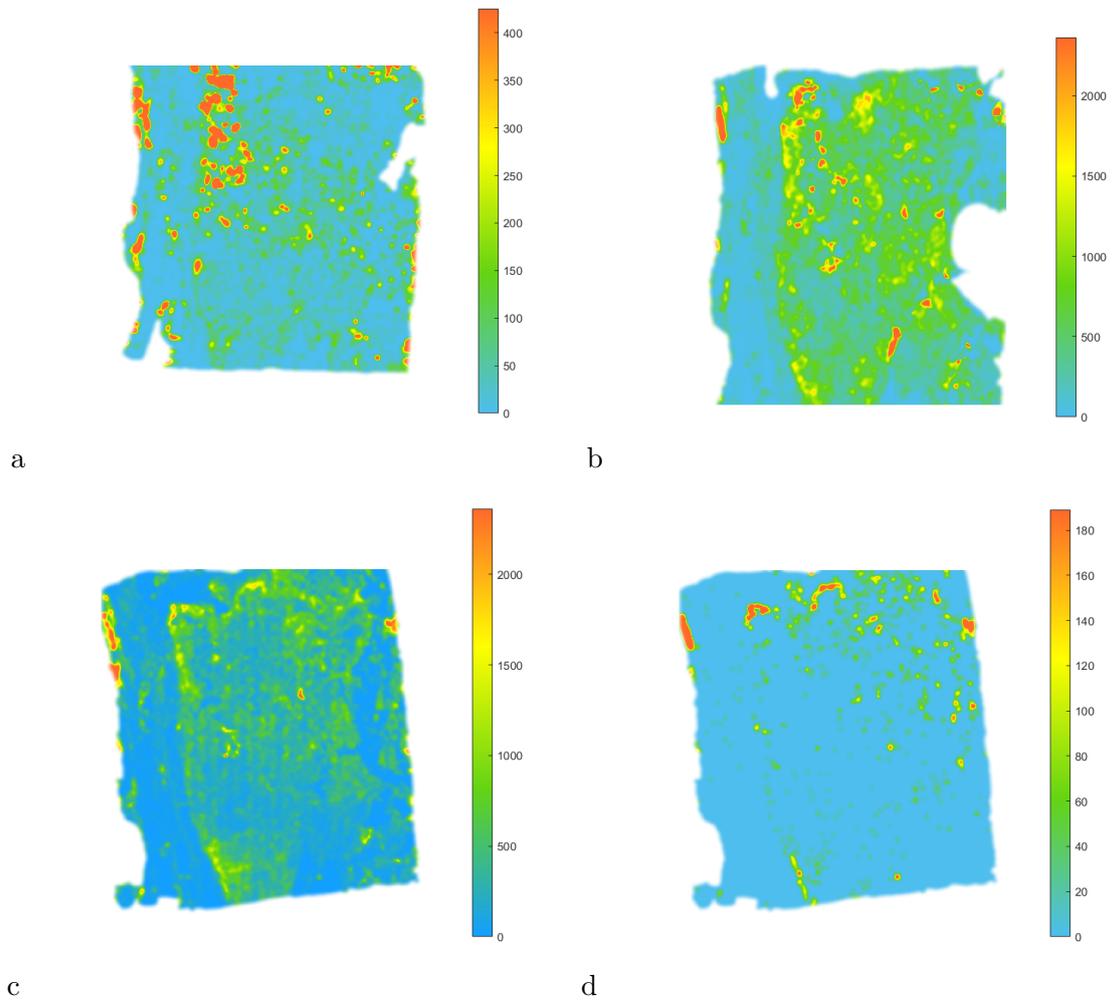


Figure C.15: Carte de densités a) CD8, b) CD163, c) CD68 cellules totales, d) CD68 grandes cellules.

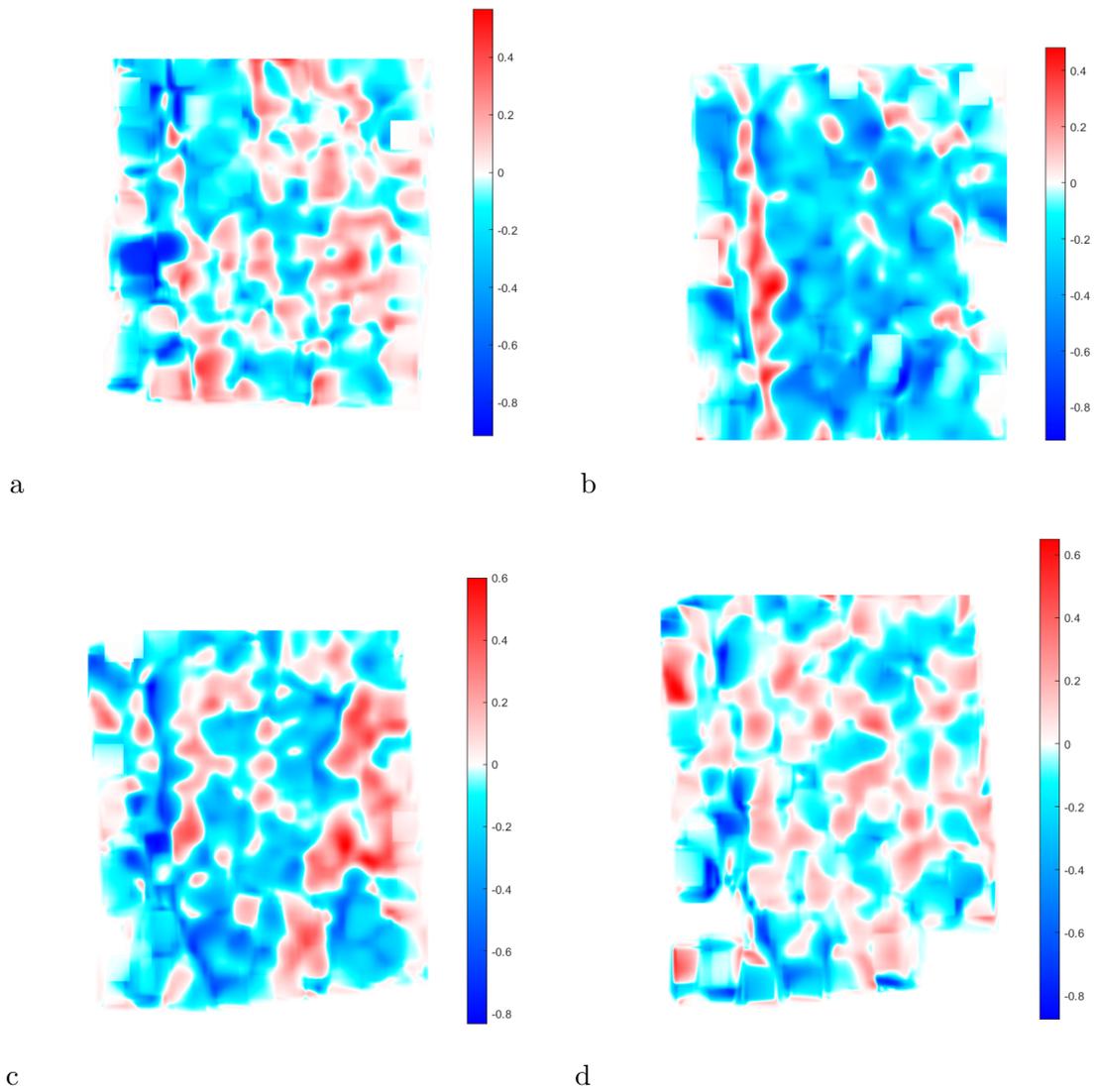


Figure C.16: Cartes de corrélations locales **a)** CD8, **b)** CD163, **c)** CD68 cellules totales, **d)** CD68 grandes cellules.

corrélation	CD8	CD163	CD68 t	CD68 g
Porosité	-0.267	-0.311	-0.113	-0.059

Table C.3: Corrélation entre porosité et densité cellulaire, pour le CD68 : t = toutes cellules marquées, g = grandes cellules marquées

C.2.3 Patient 6

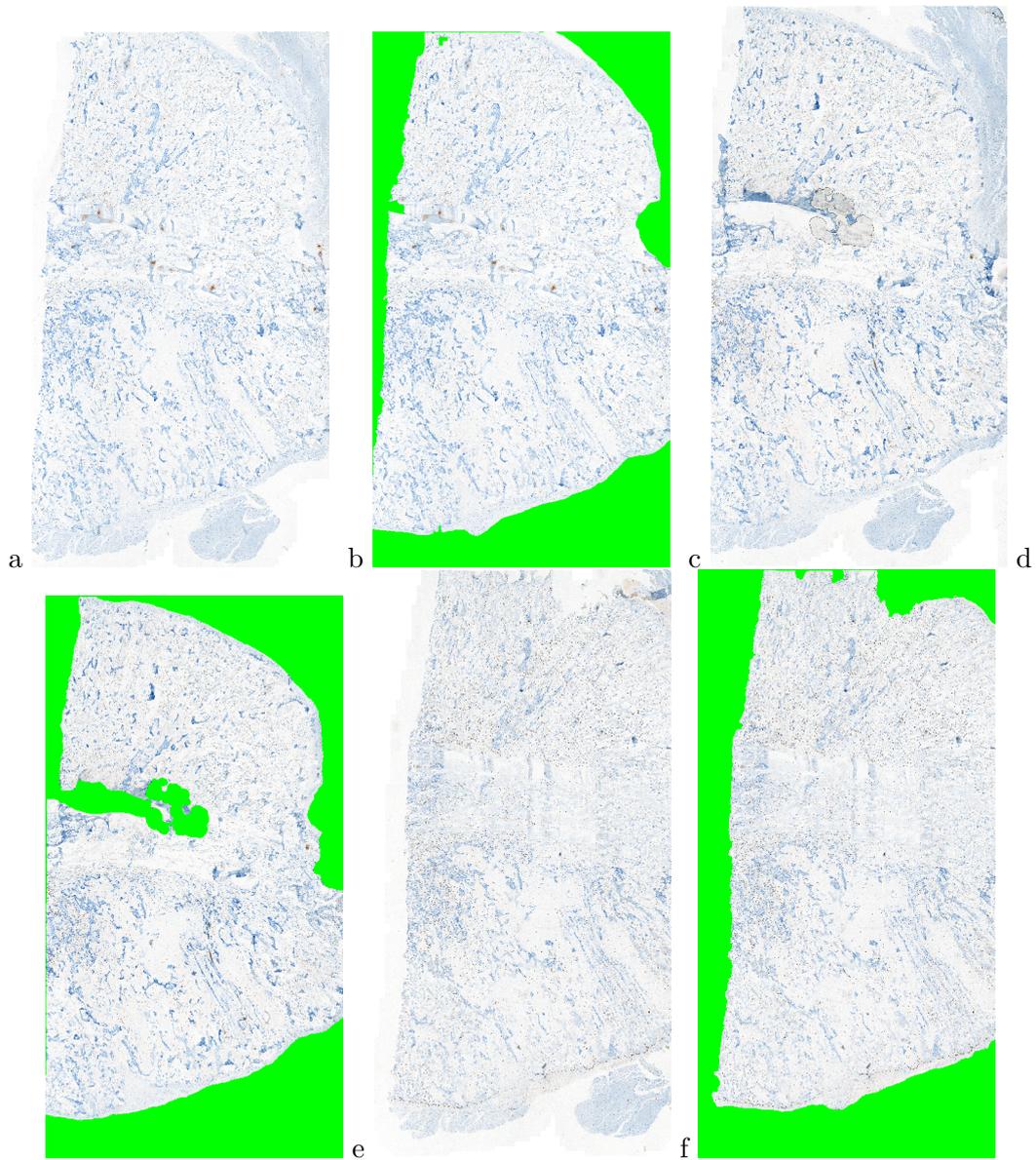


Figure C.17: Coupe immuno-histologique du patient 20T, a) marquage CD8, b) CD8 masque, c) CD163, d) CD163 masque, e) CD68, f) CD68 masque.

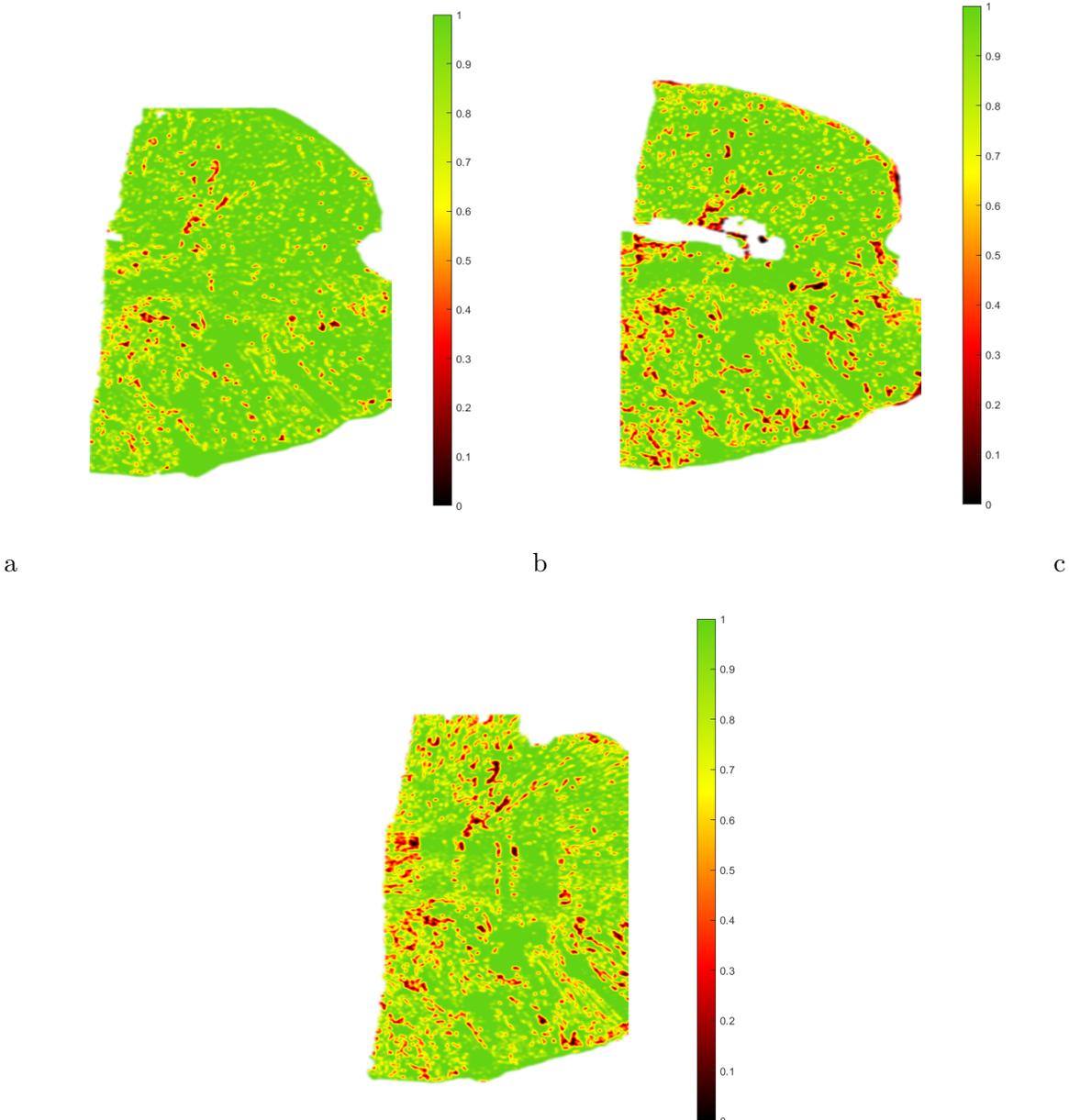


Figure C.18: Carte de porosité a) CD8, b) CD163, c) CD68.

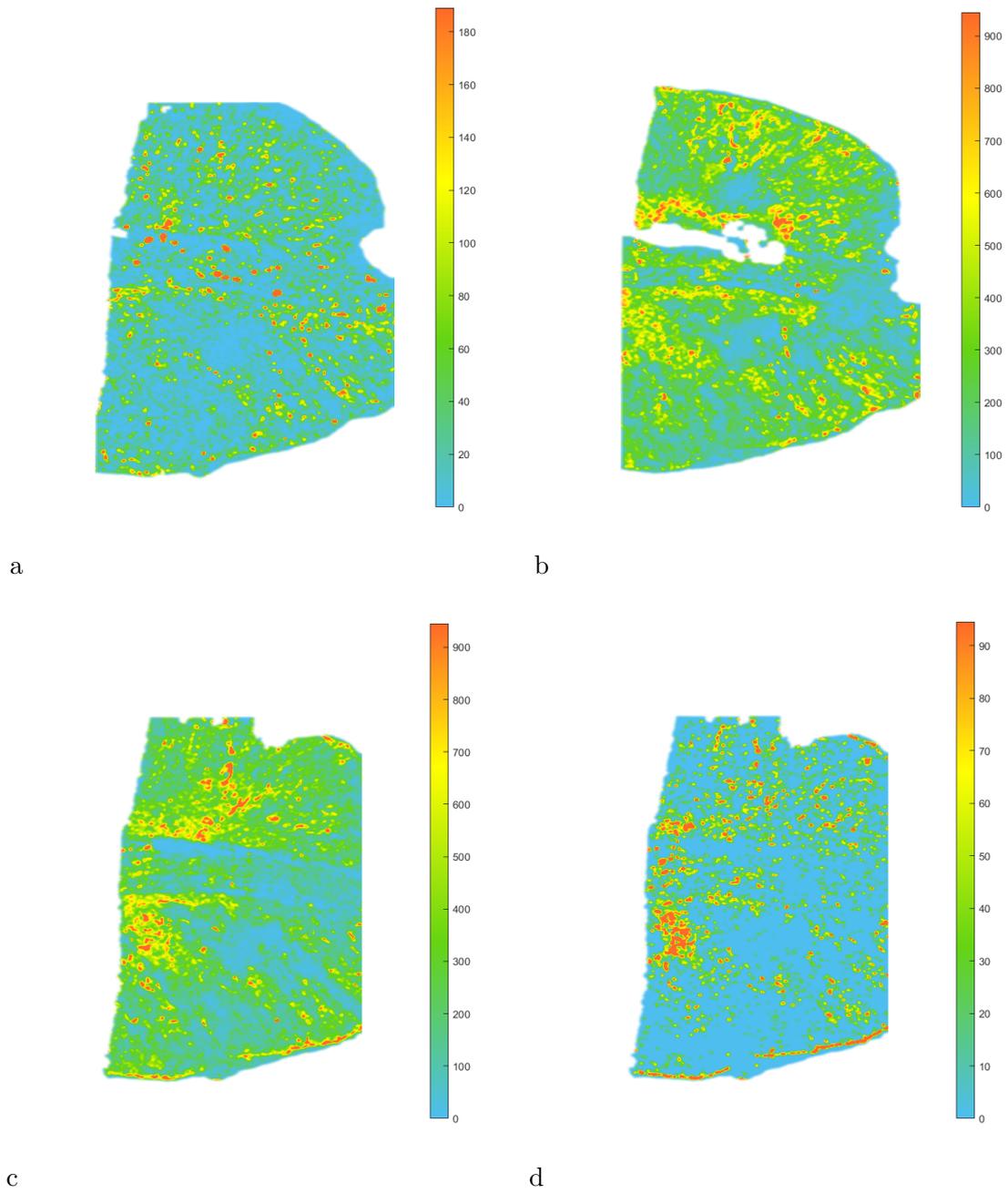


Figure C.19: Carte de densités **a)** CD8, **b)** CD163, **c)** CD68 cellules totales, **d)** CD68 grandes cellules.

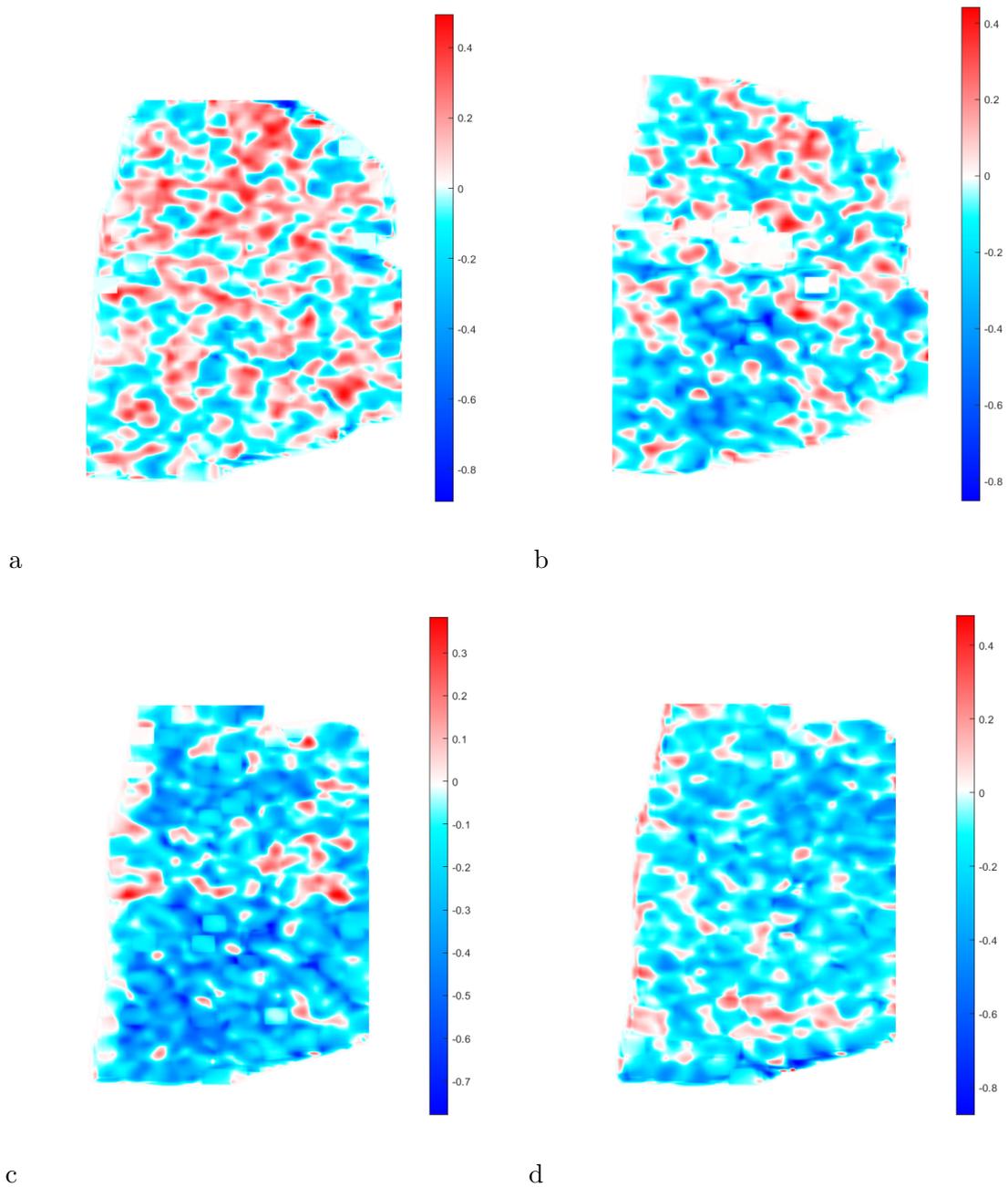


Figure C.20: Cartes de corrélations locales a) CD8, b) CD163, c) CD68 cellules totales, d) CD68 grandes cellules.

corrélation	CD8	CD163	CD68 t	CD68 g
Porosité	0.009	-0.096	-0.332	-0.249

Table C.4: Corrélacion entre porosité et densité cellulaire, pour le CD68 : t = toutes cellules marquées, g = grandes cellules marquées

C.2.4 Patient 8

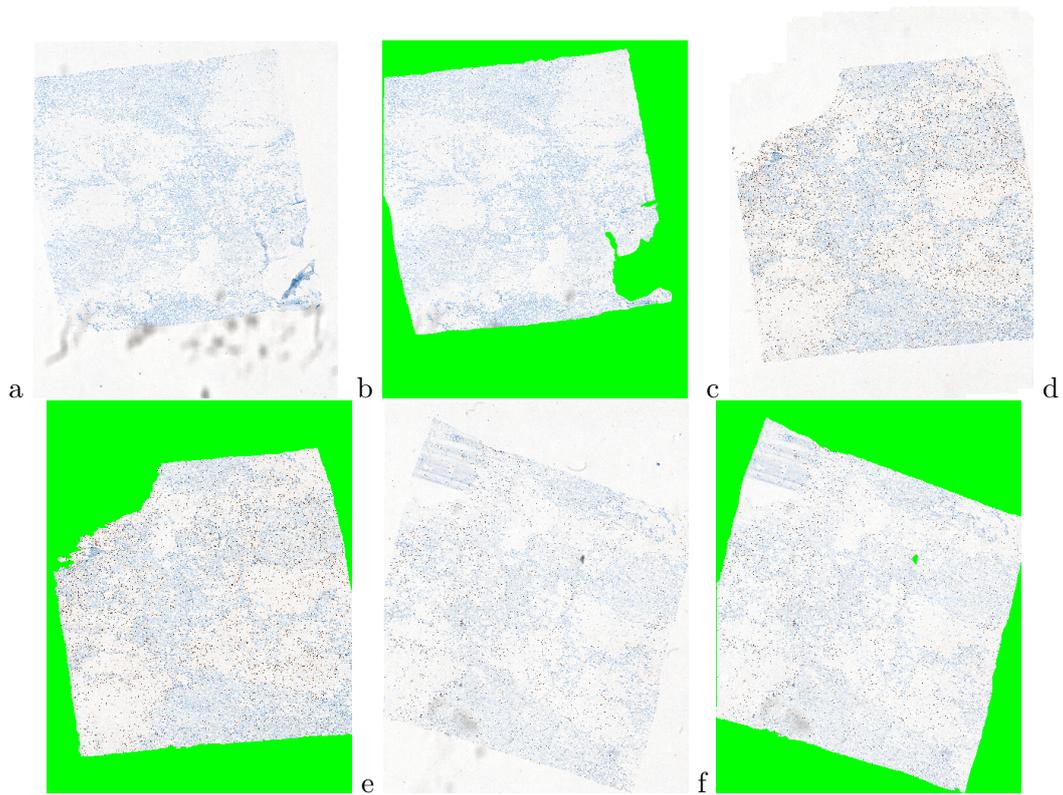


Figure C.21: Coupe immuno-histologique du patient 21TED08, a) marquage CD8, b) CD8 masque, c) CD163, d) CD163 masque, e) CD68, f) CD68 masque.

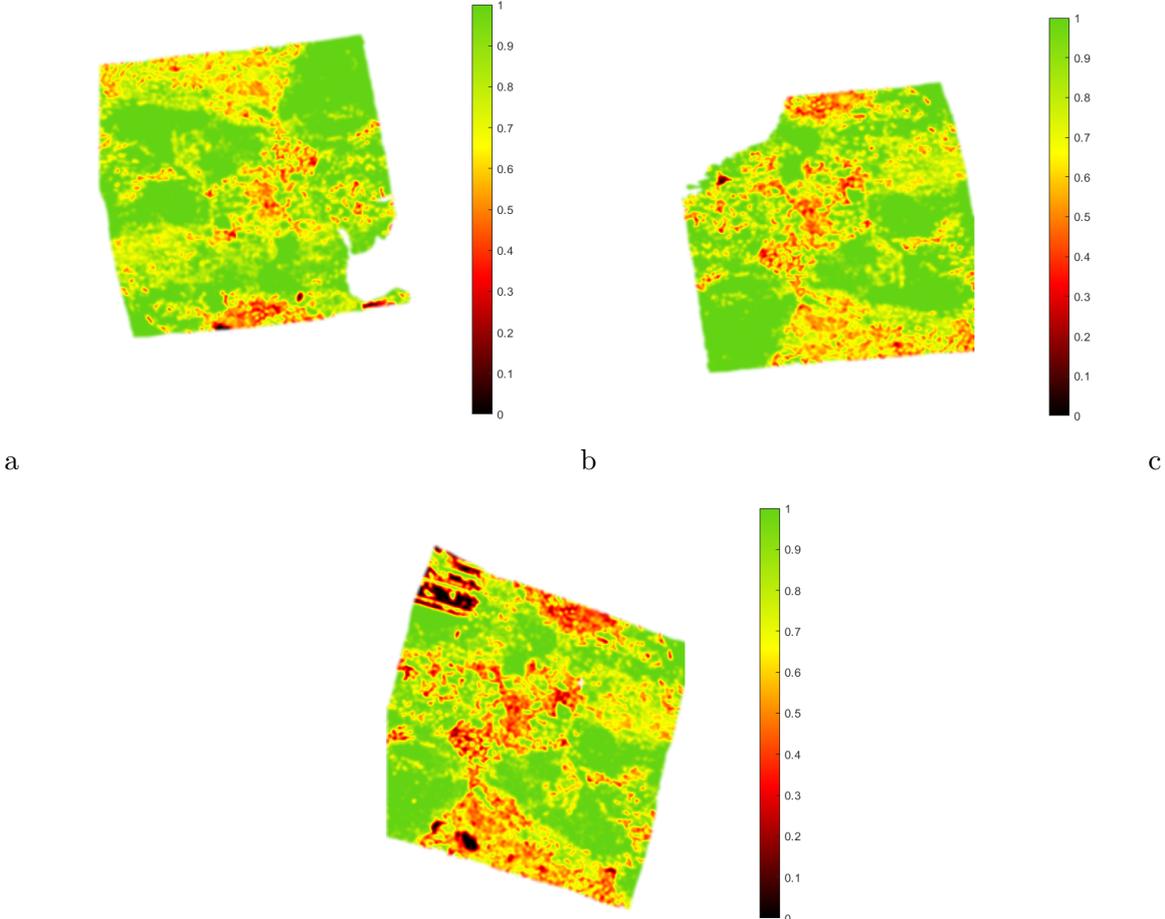


Figure C.22: Carte de porosité a) CD8, b) CD163, c) CD68.

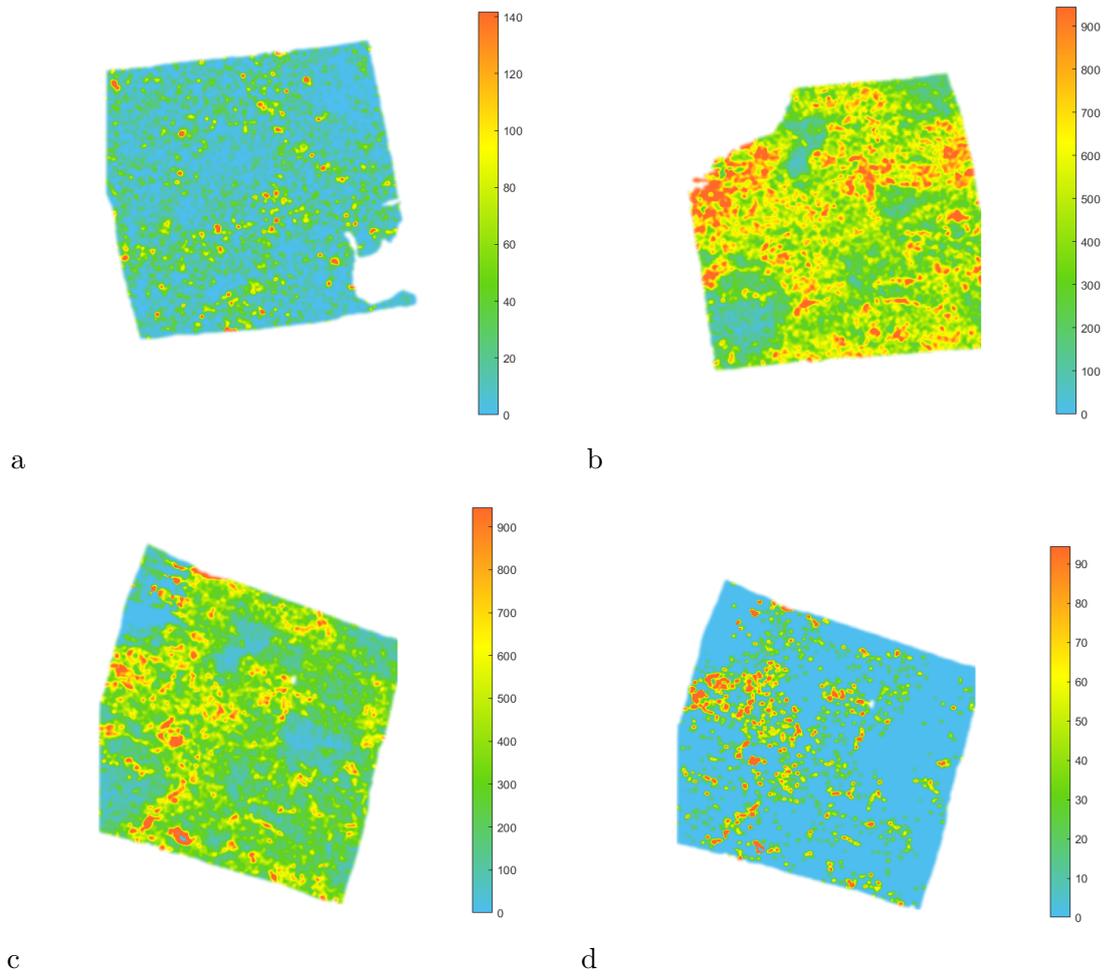


Figure C.23: Carte de densités a) CD8, b) CD163, c) CD68 cellules totales, d) CD68 grandes cellules.

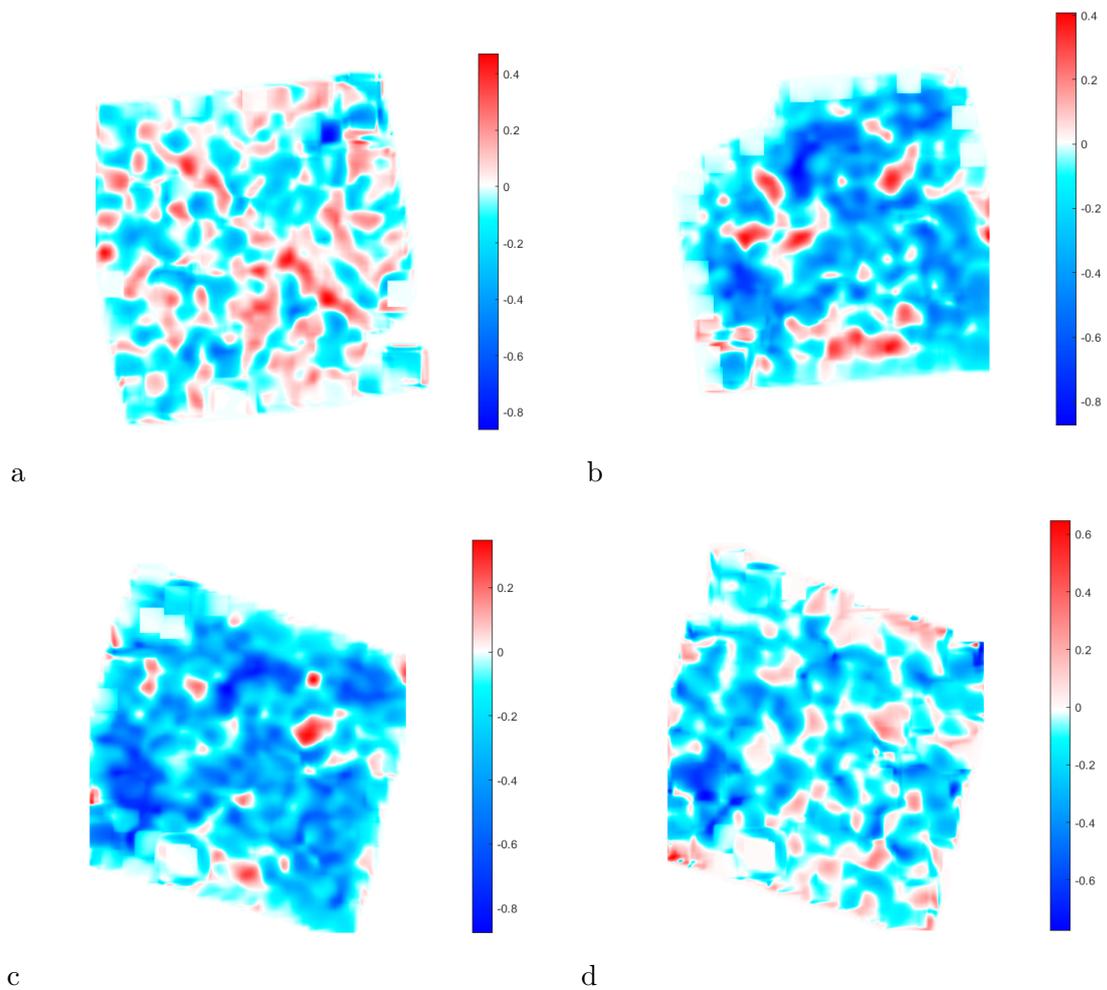


Figure C.24: Cartes de corrélations locales **a)** CD8, **b)** CD163, **c)** CD68 cellules totales, **d)** CD68 grandes cellules.

corrélacion	CD8	CD163	CD68 t	CD68 g
Porosité	-0.146	-0.408	-0.458	-0.180

Table C.5: Corrélacion entre porosité et densité cellulaire, pour le CD68 : t = toutes cellules marquées, g = grandes cellules marquées

Références

- [1] C.D.M. Fletcher, J.A. Bridge, P. Hogendoorn, and F. Mertens. *WHO Classification of Tumours of Soft Tissue and Bone. Fourth Edition. IARC WHO Classification of Tumours, No 5*. 2013.
- [2] Lisa Mirabello, Rebecca Troisi, and Sharon Savage. Osteosarcoma incidence and survival rates from 1973 to 2004 data from the surveillance, epidemiology, and end results program. *Cancer*, 115:1531–43, 02 2009.
- [3] Gaspar N, Ocean BV, Pacquement H, Bompas E, Bouvier C, Brisse HJ, Castex MP, Cheurfa N, Corradini N, Delaye J, Entz-Werlé N, Gentet JC, Italiano A, Lervat C, Marec-Berard P, Mascard E, Redini F, Saumet L, Schmitt C, Tabone MD, Verite-Goulard C, Le Deley MC, Piperno-Neumann S, and Brugieres L. Results of methotrexate-etoposide-ifosfamide based regimen (m-ei) in osteosarcoma patients included in the french os2006/sarcome-09 study. *Eur J Cancer*, 2018.
- [4] Anne Gomez-Brouchet, Eric Mascard, Aurore Siegfried, Gonzague de Pinieux, Nathalie Gaspar, Corinne Bouvier, Sébastien Aubert, Perrine Marec-Bérard, Sophie Piperno-Neumann, Béatrice Marie, Frédérique Larousserie, Christine Galant, Fabrice Fiorenza, Philippe Anract, Jérôme Sales de Gauzy, and François Guoin. Évaluation des marges de résection des sarcomes osseux traités par chimiothérapie néoadjuvante : revue de la littérature et recommandations du groupe os (groupos) du groupe sarcome français et du groupe d’étude des tumeurs osseuses (gsf-geto/resos). *Revue de Chirurgie Orthopédique et Traumatologique*, 105(4):518–527, 2019.
- [5] Xiang Chen, Armita Bahrami, Alberto Pappo, John Easton, James Dalton, Erin Hedlund, David Ellison, Sheila Shurtleff, Gang Wu, Lei Wei, Matthew Parker, Michael Rusch, Panduka Nagahawatte, Jianrong Wu, Shenghua Mao, Kristy Boggs, Heather Mulder, Donald Yergeau, Charles Lu, and Michael Dyer. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell reports*, 7:104–112, 04 2014.
- [6] Bousquet M, Noirot C, Accadbled F, Sales de Gauzy J, Castex MP, Brousset P, and Gomez-Brouchet A. Whole-exome sequencing in osteosarcoma reveals important heterogeneity of genetic alterations. 2016.
- [7] Perry JA, Kiezun A, Tonzi P, Van Allen EM, Carter SL, Baca SC, Cowley GS, Bhatt AS, Rheinbay E, Peadarallu CS, Helman E, Taylor-Weiner A, McKenna A, DeLuca DS, Lawrence MS, Ambrogio L, Sougnez C, Sivachenko A, Walensky LD, Wagle N, Mora J, de Torres C, Lavarino C, Dos Santos Aguiar S, Yunes JA, Brandalise SR, Mercado-Celis GE, Melendez-Zajgla J, Cárdenas-Cardós R, Velasco-Hidalgo L, Roberts CW, Garraway LA, Rodriguez-Galindo C, Gabriel SB, Lander ES, Golub TR, Orkin SH, Getz G, and Janeway KA. Complementary genomic approaches highlight the pi3k/mTOR pathway as a common vulnerability in osteosarcoma. 2014.
- [8] Zhou Y, Yang D, Yang Q, Lv X, Huang W, Zhou Z, Wang Y, Zhang Z, Yuan T, Ding X, Tang L, Zhang J, Yin J, Huang Y, Yu W, Wang Y, Zhou C, Su Y, He A, Sun Y, Shen Z, Qian B, Meng W, Fei J, Yao Y, Pan X, Chen P, and Hu H. Single-cell rna landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma. *Nature Communications*, 11:6322, 2020.

- [9] Feleke M, Feng W, Rothzerg E, Song D, Wei Q, Kōks S, Wood D, Liu Y, and Xu J. Single-cell rna-seq identification of four differentially expressed survival-related genes by a target: Osteosarcoma database analysis. *Exp Biol Med (Maywood)*, 2022.
- [10] Bhandari V, Hoey C, Liu LY, Lalonde E, Ray J, Livingstone J, Lesurf R, Shiah YJ, Vujcic T, Huang X, Espiritu SMG, Heisler LE, Yousif F, Huang V, Yamaguchi TN, Yao CQ, Sabelnykova VY, Fraser M, Chua MLK, van der Kwast T, Liu SK, Boutros PC, and Bristow RG. Molecular landmarks of tumor hypoxia across cancer types. *Nat Genet*, 2019.
- [11] Anne Gomez-Brouchet, Julia Gilhodes, Nathalie Van Acker, Regis Brion, Corinne Bouvier, Pauline Assemat, Nathalie Gaspar, Sebastien Aubert, Jean-Marc Guinebretiere, Beatrice Marie, Frederique Larousserie, Natacha Entz-Werlé, Gonzague De Pinieux, Eric Mascard, Francois Gouin, Pierre Brousset, Marie-Dominique Tabone, Marta Jimenez, Marie-Cecile Le Deley, Jean-Yves Blay, Laurence Brugieres, Sophie Piperno-Neumann, and Francoise Rédini. Characterization of macrophages and osteoclasts in the osteosarcoma tumor microenvironment at diagnosis: New perspective for osteosarcoma treatment? *Cancers*, 13(3), 2021.
- [12] Marchais A, Marques da Costa ME, Job B, Abbas R, Drubay D, Piperno-Neumann S, Fromigué O, Gomez-Brouchet A, Redini F, Droit R, Lervat C, Entze-Werle N, Pacquement H, Devoldere C, Cupissol D, Bodet D, Gandemer V, Berger M, Marec-Berard P, Jimenez M, Vassal G, Georger B, Brugières L, and Gaspar N. Immune infiltrate and tumor microenvironment transcriptional programs stratify pediatric osteosarcoma into prognostic groups at diagnosis. *Cancer*, 2022.
- [13] Ory B, Blanchard F, Battaglia S, Gouin F, Rédini F, and Heymann D. Zoledronic acid activates the dna s-phase checkpoint and induces osteosarcoma cell death characterized by apoptosis-inducing factor and endonuclease-g translocation independently of p53 and retinoblastoma status. *Mol Pharmacol*, 2007.
- [14] Nazon C, Pierrevelcin M, Willaume T, Lhermitte B, Weingertner N, Marco AD, Bund L, Vincent F, Bierry G, Gomez-Brouchet A, Redini F, Gaspar N, Dontenwill M, and Entz-Werle N. Together intra-tumor hypoxia and macrophagic immunity are driven worst outcome in pediatric high-grade osteosarcomas. *Cancer (Basel)*, 2022.
- [15] Gomez-Brouchet A, Illac C, Ledoux A, Fortin PY, de Barros S, Vabre C, Despas F, Peries S, Casaroli C, Bouvier C, Aubert S, de Pinieux G, Larousserie F, Galmiche L, Talmont F, Pitson S, Maddelein ML, and Cuvillier O. Sphingosine kinase-1 is overexpressed and correlates with hypoxia in osteosarcoma: Relationship with clinicopathological parameters. *Cancer (Basel)*, 2022.
- [16] Gomez-Brouchet Anne, Claire Illac, Julia Gilhodes, Corinne Bouvier, Sébastien Aubert, Guinebretière Jean-Marc, Béatrice Marie, Frédérique Larousserie, Natacha Entz-Werle, Gonzague de Pinieux, Thomas Filleron, Véronique Minard, Vincent Minville, Eric Mascard, François Gouin, Marta Jimenez, Marie-Cécile Ledelely, Sophie Piperno-Neumann, and Françoise Rédini. Cd163-positive tumor-associated macrophages and cd8-positive cytotoxic lymphocytes are powerful diagnostic markers for the therapeutic stratification of osteosarcoma patients: An immunohistochemical analysis of the biopsies from the french os2006 phase 3 trial. *OncoImmunology*, 6:00–00, 08 2017.
- [17] Pengfei Lu, Valerie M Weaver, and Zena Werb. The extracellular matrix: a dynamic niche in cancer progression. *Journal of cell biology*, 196(4):395–406, 2012.
- [18] Michael W. Pickup, Janna K. Mouw, and Valerie M. Weaver. The extracellular matrix modulates the hallmarks of cancer. *EMBO reports*, 15:1243 – 1253, 2014.
- [19] Paolo Netti, David Berk, Melody Swartz, Alan Grodzinsky, and R Jain. Role of extracellular matrix assembly in interstitial transport in solid tumors. *cancer res* 60: 2497-2503. *Cancer research*, 60:2497–503, 06 2000.

- [20] Erik Henke, Rajender Nandigama, and Süleyman Ergün. Extracellular matrix in the tumor microenvironment and its impact on cancer therapy. *Frontiers in Molecular Biosciences*, 6, 01 2020.
- [21] Jian Tian Hong-bo He Wei Luo Yu-sheng Li, Qing Liu. Angiogenesis process in osteosarcoma: An updated perspective of pathophysiology and therapeutics. *The American Journal of the Medical Sciences*, 357(4):280–288, 2019.
- [22] Franck Verrecchia and Françoise Rédini. Transforming growth factor- β signaling plays a pivotal role in the interplay between osteosarcoma cells and their microenvironment. *Frontiers in Oncology*, 8:133, 04 2018.
- [23] Richert Iseulys and Aurélie Dutour. *The immune environment of bone sarcomas*, pages 189–203. 01 2022.
- [24] T.S. Deisboeck, Z. Wang, P. Macklin, and V. Cristini. Multiscale cancer modeling. *Annual Review of Biomedical Engineering*, vol. 13, no.1, pp. 127–155, 2011.
- [25] O’Connor, James PB, Rose, Chris J, Waterton, John C, Carano, Richard AD, Parker, Geoff JM, Jackson, and Alan. Imaging intratumor heterogeneity: Role in therapy response, resistance, and clinical outcome imaging intratumor heterogeneity. *Clinical Cancer Research*, 21(2):249–257, 2015.
- [26] Andreas Heindl, Sidra Nawaz, and Yinyin Yuan. Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. *Laboratory investigation*, 95(4):377–384, 2015.
- [27] Sidra Nawaz and Yinyin Yuan. Computational pathology: Exploring the spatial dimension of tumor ecology. *Cancer letters*, 380(1):296–303, 2016.
- [28] Luong Nguyen, Akif Burak Tosun, Jeffrey L Fine, Adrian V Lee, D Lansing Taylor, and S Chakra Chennubhotla. Spatial statistics for segmenting histological structures in h&e stained tissue images. *IEEE transactions on medical imaging*, 36(7):1522–1532, 2017.
- [29] Yinyin Yuan. Spatial heterogeneity in the tumor microenvironment. *Cold Spring Harbor perspectives in medicine*, 6(8):a026583, 2016.
- [30] Rumana Rashid, Yu-An Chen, John Hoffer, Jeremy L Muhlich, Jia-Ren Lin, Robert Krueger, Hanspeter Pfister, Richard Mitchell, Sandro Santagata, and Peter K Sorger. Narrative online guides for the interpretation of digital-pathology images and tissue-atlas data. *Nature Biomedical Engineering*, 6(5):515–526, 2022.
- [31] Vipul Baxi, Robin Edwards, Michael Montalto, and Saurabh Saha. Digital pathology and artificial intelligence in translational medicine and clinical practice. *Modern Pathology*, 35(1):23–32, 2022.
- [32] Muhammad Khalid Khan Niazi, Anil V Parwani, and Metin N Gurcan. Digital pathology and artificial intelligence. *The lancet oncology*, 20(5):e253–e261, 2019.
- [33] Jialing Zhang, Stephan Stanislaw Späth, Sadie L Marjani, Wengeng Zhang, and Xinghua Pan. Characterization of cancer genomic heterogeneity by next-generation sequencing advances precision medicine in cancer treatment. *Precision clinical medicine*, 1(1):29–48, 2018.
- [34] Melissa R Junttila and Frederic J De Sauvage. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*, 501(7467):346–354, 2013.
- [35] Anne Gomez-Brouchet, Julia Gilhodes, Nathalie Van Acker, Regis Brion, Corinne Bouvier, Pauline Assemat, Nathalie Gaspar, Sébastien Aubert, Jean-Marc Guinebretiere, Béatrice Marie, et al. Characterization of macrophages and osteoclasts in the osteosarcoma tumor microenvironment at diagnosis: New perspective for osteosarcoma treatment? *Cancers*, 13(3):423, 2021.

- [36] Andrea Behanova, Anna Klemm, and Carolina Wählby. Spatial statistics for understanding tissue organization. *Frontiers in Physiology*, page 37, 2022.
- [37] Julienne L Carstens, Pedro Correa de Sampaio, Dalu Yang, Souptik Barua, Huamin Wang, Arvind Rao, James P Allison, Valerie S LeBleu, and Raghu Kalluri. Spatial computation of intratumoral t cells correlates with survival of patients with pancreatic cancer. *Nature communications*, 8(1):1–13, 2017.
- [38] Chang Gong, Robert A Anders, Qingfeng Zhu, Janis M Taube, Benjamin Green, Wenting Cheng, Imke H Bartelink, Paolo Vicini, Bing Wang, and Aleksander S Popel. Quantitative characterization of cd8+ t cell clustering and spatial heterogeneity in solid tumors. *Frontiers in oncology*, 8:649, 2019.
- [39] Joel Saltz, Rajarsi Gupta, Le Hou, Tahsin Kurc, Pankaj Singh, Vu Nguyen, Dimitris Samaras, Kenneth R Shroyer, Tianhao Zhao, Rebecca Batiste, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell reports*, 23(1):181–193, 2018.
- [40] Haoyang Mi, Chang Gong, Jeremias Sulam, Elana J Fertig, Alexander S Szalay, Elizabeth M Jaffee, Vered Stearns, Leisha A Emens, Ashley M Cimino-Mathews, and Aleksander S Popel. Digital pathology analysis quantifies spatial heterogeneity of cd3, cd4, cd8, cd20, and foxp3 immune markers in triple-negative breast cancer. *Frontiers in physiology*, 11:583333, 2020.
- [41] Henrik Failmezger, Harald Hessel, Ansh Kapil, Günter Schmidt, and Nathalie Harder. Spatial heterogeneity of cancer associated protein expression in immunohistochemically stained images as an improved prognostic biomarker. *Frontiers in Oncology*, 12, 2022.
- [42] Peter J Diggle. *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press, 2013.
- [43] Andre G Journel and Charles J Huijbregts. *Mining geostatistics*. 1976.
- [44] Jean Serra. *Image analysis and mathematical morphology. (No Title)*, 1982.
- [45] Bruce B Trangmar, Russel S Yost, and Goro Uehara. Application of geostatistics to spatial studies of soil properties. *Advances in agronomy*, 38:45–94, 1986.
- [46] Pierre Goovaerts. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biology and Fertility of soils*, 27(4):315–334, 1998.
- [47] Louis J Durlofsky. Numerical calculation of equivalent grid block permeability tensors for heterogeneous porous media. *Water resources research*, 27(5):699–708, 1991.
- [48] Yohan Davit, Christopher G Bell, Helen M Byrne, Lloyd AC Chapman, Laura S Kimpton, Georgina E Lang, Katherine HL Leonard, James M Oliver, Natalie C Pearson, Rebecca J Shipley, et al. Homogenization via formal multiscale asymptotics and volume averaging: How do the two techniques compare? *Advances in Water Resources*, 62:178–206, 2013.
- [49] Thierry Colin, Francois Cornelis, Patricio Cumsille, Guillaume Lefebvre, Clair Poignard, and Olivier Saut. Spatial modeling of tumor drug resistance: the case of liver gist metastases. *VALPARAISO'S MATHEMATICS AND ITS APPLICATIONS DAYS*, page 9, 2014.
- [50] Priti Bansal, Kshitiz Gehlot, Abhishek Singhal, and Abhishek Gupta. Automatic detection of osteosarcoma based on integrated features and feature selection using binary arithmetic optimization algorithm. *Multimedia Tools and Applications*, 81(6):8807–8834, 2022.
- [51] Anne E Carpenter, Thouis R Jones, Michael R Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A Guertin, Joo Han Chang, Robert A Lindquist, Jason Moffat, et al. Cell-profiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10):1–11, 2006.

- [52] David R Stirling, Madison J Swain-Bowden, Alice M Lucas, Anne E Carpenter, Beth A Cimini, and Allen Goodman. Cellprofiler 4: improvements in speed, utility and usability. *BMC bioinformatics*, 22(1):1–11, 2021.
- [53] Peter Bankhead, Maurice B Loughrey, José A Fernández, Yvonne Dombrowski, Darragh G McArt, Philip D Dunne, Stephen McQuaid, Ronan T Gray, Liam J Murray, Helen G Coleman, et al. Qupath: Open source software for digital pathology image analysis. *Scientific reports*, 7(1):1–7, 2017.
- [54] Christoph Sommer, Christoph Straehle, Ullrich Koethe, and Fred A Hamprecht. Ilastik: Interactive learning and segmentation toolkit. In *2011 IEEE international symposium on biomedical imaging: From nano to macro*, pages 230–233. IEEE, 2011.
- [55] Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N Straehle, Bernhard X Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, et al. Ilastik: interactive machine learning for (bio) image analysis. *Nature methods*, 16(12):1226–1232, 2019.
- [56] Fabrice De Chaumont, Stéphane Dallongeville, Nicolas Chenouard, Nicolas Hervé, Sorin Pop, Thomas Provoost, Vannary Meas-Yedid, Praveen Pankajakshan, Timothée Lecomte, Yoann Le Montagner, et al. Icy: an open bioimage informatics platform for extended reproducible research. *Nature methods*, 9(7):690–696, 2012.
- [57] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning, Data Mining, Inference, and Prediction, Second Edition*. 2009.
- [58] G. Cases A. Gomez-brouchet. Etude de la distribution spatiale du microenvironnement immunitaire dans les ostéosarcomes et corrélations aux données de l’imagerie. *Médecine spécialisée, Université Toulouse III - Paul Sabatier*, 2021.
- [59] Anne Gomez-Brouchet, Camille Franchet, Marie Cortèse, Michèle Gomez, Gisèle Viguié, Serge Estaque, and Yvan Nicaise. Atelier médico-technique tissu osseux : Aspects techniques, mégablocs en pathologie osseuse et intérêt pratique. *Bulletin de la Division Française de l’AIP*, 63:147–150, 2016.
- [60] P.G. Casali, S. Bielack, N. Abecassis, H.T. Aro, S. Bauer, R. Biagini, S. Bonvalot, I. Boukovinas, J V M G Bovee, B. Brennan, and et al. Bone sarcomas: Esmo–paedcan–euracan clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 29:iv79–iv95, 2018.
- [61] Frédéric Devernay. Cours détection de contours. <http://devernay.free.fr/cours/vision/pdf/c3.pdf>.
- [62] Kwang In Kim, Keechul Jung, Se Hyun Park, and Hang Joon Kim. Support vector machines for texture classification. *IEEE transactions on pattern analysis and machine intelligence*, 24(11):1542–1550, 2002.
- [63] Vishal Monga, Yuelong Li, and Yonina C Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021.
- [64] Jean-Luc Starck, Jalal Fadili, and Fionn Murtagh. The undecimated wavelet decomposition and its reconstruction. *IEEE transactions on image processing*, 16(2):297–309, 2007.
- [65] Anna Linderhed. Image empirical mode decomposition: A new tool for image processing. *Advances in Adaptive Data Analysis*, 1(02):265–294, 2009.
- [66] Roberto H Bamberg and Mark JT Smith. A filter bank for the directional decomposition of images: Theory and design. *IEEE transactions on signal processing*, 40(4):882–893, 1992.

- [67] Jean-François Aujol, Guy Gilboa, Tony Chan, and Stanley Osher. Structure-texture image decomposition—modeling, algorithms, and parameter selection. *International journal of computer vision*, 67:111–136, 2006.
- [68] Kartic Subr, Cyril Soler, and Frédo Durand. Edge-preserving multiscale image decomposition based on local extrema. *ACM Transactions on Graphics (TOG)*, 28(5):1–9, 2009.
- [69] Norbert Wiener. Extrapolation, interpolation, and smoothing of stationary time series: With engineering applications. 1949.
- [70] Hidemitsu Ogawa and Erkki Oja. Projection filter, wiener filter, and karhunen-loeve subspaces in digital image restoration. *Journal of Mathematical analysis and applications*, 114(1):37–51, 1986.
- [71] Yohan Davit, Christopher G. Bell, Helen M. Byrne, Lloyd A.C. Chapman, Laura S. Kimp-ton, Georgina E. Lang, Katherine H.L. Leonard, James M. Oliver, Natalie C. Pearson, Rebecca J. Shipley, Sarah L. Waters, Jonathan P. Whiteley, Brian D. Wood, and Michel Quintard. Homogenization via formal multiscale asymptotics and volume averaging: How do the two techniques compare? *Advances in Water Resources*, 62:178 – 206, 2013. A tribute to Stephen Whitaker.
- [72] Geoff Bohling. Introduction to geostatistics and variogram analysis. *Kansas geological survey*, 1:1–20, 2005.
- [73] Wafa Abid, Khaled Taouil, and Mohamed Salim Bouhleh. Segmentation des coupes histologiques osseuses par la ligne de partage des eaux. In *4th International Conference: Sciences Of Electronic, Technologies Of Information And Telecommunications*, 2007.
- [74] Stephen C. Cowin. Bone poroelasticity. *Journal of Biomechanics*, 32(3):217–238, 1999.
- [75] Dimitrios Hadjidakis and Ioannis Androulakis. Bone remodeling. *Annals of the New York Academy of Sciences*, 1092:385–96, 01 2007.
- [76] Isabel F Tresguerres, Miguel Alobera, Mariano Pingarron, and Luis Blanco. Physiological bases of bone regeneration ii. the remodeling process. *Medicina oral, patología oral y cirugía bucal*, 11:E151–7, 04 2006.
- [77] Charles H. Turner. Three rules for bone adaptation to mechanical stimuli. *Bone*, 23 5:399–407, 1998.
- [78] Iain Kalfas. Principles of bone healing. *Neurosurgical focus*, 10:E1, 02 2001.
- [79] Elisabeth H. Burger, Jenneke Klein-Nulend, and Stephen C. Cowin. Mechanotransduction in bone. In *Molecular and Cellular Biology of Bone*, volume 5 of *Advances in Organ Biology*, pages 123–136. Elsevier, 1998.
- [80] Christopher Jacobs, Sara Temiyasathit, and Alesha Castillo. *Osteocyte Mechanobiology and Pericellular Mechanics*, volume 12, pages 369–400. 08 2010.
- [81] Lei Qin, Wen Liu, Huiling Cao, and Guozhi Xiao. Molecular mechanosensors in osteocytes. *Bone Research*, 8, 12 2020.
- [82] Claudia Wittkowske, Gwendolen Reilly, Damien Lacroix, and Cecile Perrault. In vitro bone cell models: Impact of fluid shear stress on bone formation. *Frontiers in Bioengineering and Biotechnology*, 4, 11 2016.
- [83] Adrian Shieh. Biomechanical forces shape the tumor microenvironment. *Annals of biomedical engineering*, 39:1379–89, 05 2011.
- [84] Yi-Xian Qin, Clinton Rubin, and Kenneth McLeod. Nonlinear dependence of loading intensity and cycle number in the maintenance of bone mass and morphology. *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*, 16:482–9, 07 1998.

- [85] Andrew Ahn and Alan Grodzinsky. Relevance of collagen piezoelectricity to "wolff's law": A critical review. *Medical engineering & physics*, 31:733–41, 04 2009.
- [86] Steinar Evje and Jahn Waldeland. How tumor cells possibly can make use of interstitial fluid flow in a strategy for metastasis. *Cellular and Molecular Bioengineering*, 03 2019.
- [87] Claudia Mierke. The fundamental role of mechanical properties in the progression of cancer disease and inflammation. *Reports on Progress in Physics*, 77:076602, 07 2014.
- [88] Letícia E Charelli, João PD Ferreira, Carolina P Naveira-Cotta, and Tiago A Balbino. Engineering mechanobiology through organoids-on-chip: A strategy to boost therapeutics. *Journal of Tissue Engineering and Regenerative Medicine*, 15(11):883–899, 2021.
- [89] A Bel-Brunon, Michel Coret, K Bruyère-Garnier, and A Combescure. Comparison of two homogenization methods using a damage model for a fibrous membrane, based on the fibers' fracture process at the microscale. *European Journal of Mechanics-A/Solids*, 39:1–10, 2013.
- [90] Moussa Kfoury, Benoit Noetinger, Michel Quintard, and Rachid Ababou. Upscaling fractured heterogeneous media: Permeability and mass exchange coefficient. *Journal of Applied Mechanics*, 73, 01 2006.
- [91] Romain Guibert, Pierre Horgue, Debenest Gerald, and Michel Quintard. A comparison of various methods for the numerical evaluation of porous media permeability tensors from pore-scale geometry. *Mathematical geosciences*, 48:329–347, 04 2016.
- [92] Lars Holden and Oddvar Lia. A tensor estimator for the homogenization of absolute permeability. *Transport in Porous Media*, 8:37–46, 01 1992.
- [93] Martin S. Alnæs, Jan Blechta, Johan Hake, August Johansson, Benjamin Kehlet, Anders Logg, Chris Richardson, Johannes Ring, Marie E. Rognes, and Garth N. Wells. The fenics project version 1.5. *Archive of Numerical Software*, 3(100), 2015.
- [94] Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-d finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79:1309 – 1331, 09 2009.
- [95] A. Henderson. Paraview guide, a parallel visualization application. *Kitware Inc.*, 2007.
- [96] Stephen Whitaker. *The Method of Volume Averaging*. 01 1999.
- [97] Ching-Wei Wang, Eric Budiman Gosno, and Yen-Sheng Li. Fully automatic and robust 3d registration of serial-section microscopic images. *Scientific reports*, 5(1):1–14, 2015.